Vol. 63, No. 5 September, 1956

Psychological Review

THEODORE M. NEWCOMB, Editor
University of Michigan
Lorraine Bouthilet, Managing Editor

CONTENTS

PUBLISHED BIMONTHLY BY THE
AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.

The *Psychological Review* is devoted to theoretical articles of significance to any area of psychology. Except for occasional articles solicited by the Editor, manuscripts exceeding twelve printed pages (about 7,500 words) are not accepted. Ordinarily manuscripts which consist primarily of original reports of research should be submitted to other journals.

Because of the large number of manuscripts submitted, there is an inevitable publication lag of several months. Authors may avoid this delay if they are prepared to pay the costs of publishing their own articles; the appearance of articles by other contributors is not thereby delayed.

Tables, footnotes, and references should appear on separate pages; all of these, as well as the text, should be typed double-spaced throughout, in all manuscripts submitted. All manuscripts should be submitted in duplicate. Original figures are prepared for publication; duplicate figures may be photographic or pencil-drawn copies. Authors are cautioned to retain a copy of the manuscript to guard against loss in the mail. Manuscripts should be addressed to the Editor, Dr. Theodore M. Newcomb, Doctoral Program in Social Psychology, University of Michigan, Ann Arbor, Michigan.

PUBLISHED BIMONTHLY BY THE

AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.

PRINCE AND LEMON STREETS, LANCASTER, PA.

1333 SIXTEENTH ST. N. W., WASHINGTON 6, D. C.

$8.00 volume $1.50 issue

Entered as second-class matter July 13, 1897, at the post-office at Lancaster, Pa., under Act of Congress of March 3, 1879

Acceptance for mailing at the special rate of postage provided for in paragraph (d-2), Section 34.40, P. L. & R. of 1948, authorized Jan. 8, 1948

Send all communications, including address changes, to 1333 Sixteenth St. N.W., Washington 6, D. C. Address changes must arrive by the 10th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

Vol. 63, No. 5 September, 1956

THE PSYCHOLOGICAL REVIEW

STRUCTURAL BALANCE: A GENERALIZATION OF HEIDER'S THEORY [1]

DORWIN CARTWRIGHT AND FRANK HARARY
Research Center for Group Dynamics, University of Michigan

A persistent problem of psychology has been how to deal conceptually with patterns of interdependent properties. This problem has been central, of course, in the theoretical treatment by Gestalt psychologists of phenomenal or neural *configurations* or *fields* (**12**, **13**, **15**). It has also been of concern to social psychologists and sociologists who attempt to employ concepts referring to social *systems* (**18**).

Heider (**19**), reflecting the general field-theoretical approach, has considered certain aspects of cognitive fields which contain perceived people and impersonal objects or events. His analysis focuses upon what he calls the *P-O-X* unit of a cognitive field, consisting of *P* (one person), *O* (another person), and *X* (an impersonal entity). Each relation among the parts of the unit is conceived as interdependent with each other relation. Thus, for example, if *P* has a relation of affection for *O* and if *O* is seen as responsible for *X*, then there will be a tendency for *P* to like or approve of *X*. If the nature of *X* is such that it would "normally" be evaluated as bad, the whole *P-O-X* unit is placed in a state of imbalance, and pressures will arise to change it toward a state of balance. These pressures may work to change the relation of affection between *P* and *O*, the relation of responsibility between *O* and *X*, or the relation of evaluation between *P* and *X*.

The purpose of this paper is to present and develop the consequences of a formal definition of balance which is consistent with Heider's conception and which may be employed in a more general treatment of empirical configurations. The definition is stated in terms of the mathematical theory of linear graphs (**8**, **14**) and makes use of a distinction between a given relation and its opposite relation. Some of the ramifications of this definition are then examined by means of theorems derivable from the definition and from graph theory.

Heider's Conception of Balance

In developing his analysis of balanced cognitive units, Heider distinguishes between two major *types* of relations. The first concerns attitudes, or the relation of liking or evaluating. It is represented symbolically as **L** when positive and as **∼L** when negative. Thus, *P***L***O* means *P* likes, loves, values, or approves *O*, and *P***∼L***O* means *P* dislikes, negatively

[1] This paper was prepared as part of a project sponsored in the Research Center for Group Dynamics by the Rockefeller Foundation.

values, or disapproves O. The second type of relation refers to cognitive unit formation, that is, to such specific relations as similarity, possession, causality, proximity, or belonging. It is written as **U** or $\sim$**U**. Thus, according to Heider, P**U**X means that P owns, made, is close to, or is associated with X, and $P\sim$**U**X means that P does not own, did not make, or is not associated with X.

A *balanced state* is then defined in terms of certain combinations of these relations. The definition is stated separately for two and for three entities.

> In the case of two entities, a balanced state exists if the relation between them is positive (or negative) in all respects, i.e., in regard to all meanings of **L** and **U** . . . . In the case of three entities, a balanced state exists if all three relations are positive in all respects, or if two are negative and one positive (9, p. 110).

These are examples of balanced states: P likes something he made (P**U**X, P**L**X); P likes what his friend likes (P**L**O, O**L**X, P**L**X); P dislikes what his friend dislikes (P**L**O, $O\sim$**L**X, $P\sim$**L**X); P likes what his enemy dislikes ($P\sim$**L**O, $O\sim$**L**X, P**L**X); and P's son likes what P likes (P**U**O, P**L**X, O**L**X).

Heider's basic hypothesis asserts that there is a tendency for cognitive units to achieve a balanced state. Pressures toward balance may produce various effects.

> If no balanced state exists, then forces towards this state will arise. Either the dynamic characters will change, or the unit relations will be changed through action or through cognitive reorganization. If a change is not possible, the state of imbalance will produce tension (9, pp. 107–109).

The theory, stated here in sketchy outline, has been elaborated by Heider so as to treat a fuller richness of cognitive experience than would be suggested by our brief description. It has been used, too, by a number of others as a point of departure for further theoretical and empirical work. We shall summarize briefly some of the major results of this work.

Horowitz, Lyons, and Perlmutter (**10**) attempted to demonstrate tendencies toward balance in an experiment employing members of a discussion group as subjects. At the end of a discussion period each subject was asked to indicate his evaluation of an event (P**L**X or $P\sim$**L**X) which had occurred during the course of the discussion. The event selected for evaluation was one which would be clearly seen as having been produced by a single person (O**U**X). The liking relation between each P and O (P**L**O or $P\sim$**L**O) had been determined by a sociometric questionnaire administered before the meeting. Would P's evaluation of the event be such as to produce a balanced P-O-X unit? If so, P's evaluation of O and X should be of the same sign. The experimental data tend to support the hypothesis that a P-O-X unit tends toward a balanced state.[2]

The social situation of a discussion group can be better analyzed, according to Horowitz, Lyons, and Perlmutter, by considering a somewhat more complex cognitive unit. The evaluation of X made by P, they argue, will be determined not only by P's evaluation of O but also by his perception of the evaluation of X given by others (Qs) in the group. The basic unit of such a social situation, then, consists of the subject, a

[2] One of the attractive features of this study is that it was conducted in a natural "field" setting, thus avoiding the dangers of artificiality. At the same time the setting placed certain restrictions on the possibility of manipulation and control of the variables. The data show a clear tendency for P to place a higher evaluation on Xs produced by more attractive Os. It is not clearly demonstrated that P likes Xs produced by liked Os and dislikes Xs produced by disliked Os.

person who is responsible for the event, and another person who will be seen by the subject as supporting or rejecting the event. This is called a *P-O-Q-X* unit. The additional data needed to describe these relations were obtained from the sociometric questionnaire which indicated *P*'s evaluation of *Q* (*P***L***Q* or *P*~**L***Q*), and from a question designed to reveal *P*'s perception of *Q*'s support or rejection of *X*, treated by the authors as a unit relation (*Q***U***X* or *Q*~**U***X*).[3]

Although these authors indicate the possibility of treating the *P-O-Q-X* unit in terms of balance, they do not develop a formal definition of a balanced configuration consisting of four elements. They seem to imply that the *P-O-Q-X* unit will be balanced if the *P-O-X* and the *P-Q-X* units are both balanced. They do not consider the relation between *Q* and *O*, nor the logically possible components of which it could be a part. Their analysis is concerned primarily with the two triangles (*P-O-X* and *P-Q-X*), which are interdependent, since both contain the relation of *P*'s liking of *X*. We noted above that the data tend to support the hypothesis that the *P-O-X* unit will tend toward balance. The data even more strongly support the hypothesis when applied to the *P-Q-X* unit; *P*'s evaluation of *X* and his perception of *Q*'s attitude toward *X* tend to agree when *P* likes *Q*, and to disagree when *P* dislikes *Q*. It should be noted, however, that there was also a clear tendency for *P* to see *Q*'s evaluation of *X* as agreeing with his own whether or not he likes *Q*.

In a rather different approach to the question of balanced *P-O-X* units, Jordan (**11**) presented subjects with 64 different hypothetical situations in which the **L** and **U** relations between each pair of elements was systematically varied. The subject was asked to place himself in each situation by taking the part of *P*, and to indicate on a scale the degree of pleasantness or unpleasantness he experienced. Unpleasantness was assumed to reflect the postulated tension produced by imbalanced units. Jordan's data tend to support Heider's hypothesis that imbalanced units produce a state of tension, but he too found that additional factors need to be considered. He discovered, for example, that negative relations were experienced as unpleasant even when contained in balanced units. This unpleasantness was particularly acute when *P* was a part of the negative relation. Jordan's study permits a detailed analysis of these additional influences, which we shall not consider here.

Newcomb (**17**), in his recent theory of interpersonal communication, has employed concepts rather similar to those of Heider. He conceives of the simplest communicative act as one in which one person *A* gives information to another person *B* about something *X*. The similarity of this *A-B-X* model to Heider's *P-O-X* unit, together with its applicability to objective interpersonal relations (rather than only to the cognitive structure of a single person), may be seen in the following quotations from Newcomb:

> *A-B-X* is . . . regarded as constituting a system. That is, certain definable relationships between *A* and *B*, between *A* and *X*, and between *B* and *X* are all viewed as interdependent. . . . For some purposes the system may be regarded as a phenomenal one within the life space of *A* or *B*, for other purposes as an "objective" system including all the possible relationships as inferred from observations of *A*'s and *B*'s behavior (**17**, p. 393).

[3] Whether this relation should be treated as **U** or **L** is subject to debate. For testing Heider's theory of balance, however, the issue is irrelevant, since he holds that the two relations are interchangeable in defining balance.

Newcomb then develops the concept of "strain toward symmetry," which appears to be a special instance of Heider's more general notion of "tendency toward balance." "Strain toward symmetry" is reflected in several manifestations of a tendency for *A* and *B* to have attitudes of the same sign toward a common *X*. Communication is the most common and usually the most effective manifestation of this tendency.

By use of this conception Newcomb reinterprets several studies (**1**, **4**, **5**, **16**, **20**) which have investigated the interrelations among interpersonal attraction, tendencies to communicate, pressures to uniformity of opinion among members of a group, and tendencies to reject deviates. The essential hypothesis in this analysis is stated thus:

> If *A* is free either to continue or not to continue his association with *B*, one or the other of two eventual outcomes is likely: (*a*) he achieves an equilibrium characterized by relatively great attraction toward *B* and by relatively high perceived symmetry, and the association is continued; or (*b*) he achieves an equilibrium characterized by relatively little attraction toward *B* and by relatively low perceived symmetry, and the association is discontinued (**17**, p. 402).

Newcomb's outcome *a* is clearly a balanced state as defined by Heider. Outcome *b* cannot be unambiguously translated into Heider's terms. If by "relatively little attraction toward *B*" is meant a negative **L** relation between *A* and *B*, then this outcome would also seem to be balanced. Newcomb's "continuation or discontinuation of the association between *A* and *B*" appear to correspond to Heider's **U** and **∼U** relations.

STATEMENT OF THE PROBLEM

This work indicates that the tendency toward balance is a significant determinant of cognitive organization, and that it may also be important in interpersonal relations. The concept of balance, however, has been defined so as to apply to a rather limited range of situations, and it has contained certain ambiguities. We note five specific problems.

1. *Unsymmetric relations.* Should all relations be conceived as symmetric? The answer is clearly that they should not; it is possible for *P* to like *O* while *O* dislikes *P*. In fact, Tagiuri, Blake, and Bruner (**21**) have intensively studied dyadic relations to discover conditions producing symmetric relations of actual and perceived liking. Theoretical discussions of balance have sometimes recognized this possibility—Heider, for example, states that unsymmetric liking is unbalanced—but there has been no general definition of balance which covers unsymmetric relations. The empirical studies of balance have assumed that the relations are symmetric.

2. *Units containing more than three entities.* Nearly all theorizing about balance has referred to units of three entities. While Horowitz, Lyons, and Perlmutter studied units with four entities, they did not *define* balance for such cases. It would seem desirable to be able to speak of the balance of even larger units.

3. *Negative relations.* Is the negative relation the *complement* of the relation or its *opposite?* All of the discussions of balance seem to equate these, but they seem to us to be quite different, for the complement of a relation is expressed by adding the word "not" while the opposite is indicated by the prefix "dis" or its equivalent. Thus, the complement of "liking" is "not liking"; the opposite of "liking" is "disliking." In general, it appears that **∼L** has been taken to mean "dislike" (the opposite relation) while

~**U** has been used to indicate "not associated with" (the complementary relation). Thus, for example, Jordan says: "Specifically, '+**L**' symbolizes a positive attitude, '−**L**' symbolizes a negative attitude, '+**U**' symbolizes the existence of unit formation, and '−**U**' symbolizes the lack of unit formation" (**11**, p. 274).

4. *Relations of different types.* Heider has made a distinction between two types of relations—one based upon liking and one upon unit formation. The various papers following up Heider's work have continued to use this distinction. And it seems reasonable to assume that still other types of relations might be designated. How can a definition of balance take into account relations of different types? Heider has suggested some of the ways in which liking and unit relations may be combined, but a general formulation has yet to be developed.

5. *Cognitive fields and social systems.* Heider's intention is to describe balance of cognitive units in which the entities and relations enter as experienced by a single individual. Newcomb attempts to treat social systems which may be described objectively. In principle, it should be possible also to study the balance of sociometric structures, communication networks, patterns of power, and other aspects of social systems.

We shall attempt to define balance so as to overcome these limitations. Specifically, the definition should (*a*) encompass unsymmetric relations, (*b*) hold for units consisting of any finite number of entities, (*c*) preserve the distinction between the *complement* and the *opposite* of a relation, (*d*) apply to relations of different types, and (*e*) serve to characterize cognitive units, social systems, or any configuration where both a relation and its opposite must be specified.

THE CONCEPTS OF GRAPH, DIGRAPH, AND SIGNED GRAPH

Our approach to this problem has two primary antecedents: (*a*) Lewin's treatment (**15**) of the concepts of whole, differentiation, and unity, together with Bavelas' extension (**2**) of this work to group structure; and (*b*) the mathematical theory of linear graphs.

Many of the graph-theoretic definitions given in this section are contained in the classical reference on graph theory, König (**14**), as well as in Harary and Norman (**8**). We shall discuss, however, those concepts which lead up to the theory of balance.

A ***linear graph***, or briefly a ***graph***, consists of a finite collection of ***points***[4] $A, B, C, \cdots$ together with a prescribed subset of the set of all unordered pairs of distinct points. Each of these unordered pairs, AB, is a *line* of the graph. (From the viewpoint of the theory of binary relations,[5] a graph corresponds to an irreflexive[6] symmetric relation on points $A, B, C, \cdots$. Alternatively a graph may be represented as a matrix.[7])

Figure 1 depicts a graph of four points and four lines. The points might represent people, and the lines some relationship such as mutual liking. With this interpretation, Fig. 1 indicates that mutual liking exists between those pairs of people A, B, C, and D joined by lines. Thus D is in the relation with all other persons, while C is in the relation only with D.

[4] Points are often called "vertices" by mathematicians and "nodes" by electrical engineers.

[5] This is the approach used by Heider.

[6] A relation is irreflexive if it contains no ordered pairs of the form (a, a), i.e., if no element is in this relation to itself.

[7] This treatment is discussed in Festinger (**3**). The logical equivalence of relations, graphs, and matrices is taken up in Harary and Norman (**8**).

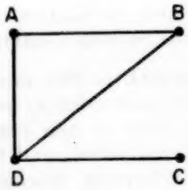


FIG. 1. A linear graph of four points and four lines. The presence of line AB indicates the existence of a specified symmetric relationship between the two entities A and B.

Figure 1 could be used, of course, to represent many other kinds of relationships between many other kinds of entities.

It is apparent from this definition of graph that relations are treated in an all-or-none manner, i.e., either a relation exists between a given pair of points or it does not. Obviously, however, many relationships of interest to psychologists (liking, for example) exist in varying degrees. This fact means that our present use of graph theory can treat only the structural, and not the numerical, aspects of relations. While our treatment is thereby an incomplete representation of the strength of relations, we believe that conceptualization of the structural properties of relations is a necessary first step toward a more adequate treatment of the more complex situations. Such an elaboration, however, goes beyond the scope of this paper.

A *directed graph*, or a *digraph*, consists of a finite collection of points together with a prescribed subset of the set of all ordered pairs of distinct points. Each of these ordered pairs $\overrightarrow{AB}$ is called a *line* of the digraph. Note that the only difference between the definitions of graph and digraph is that the lines of a graph are unordered pairs of points while the lines of a digraph are ordered pairs of points. An *ordered pair* of points is distinguished from an unordered pair by designating one of the points as the first point and the other as the second. Thus, for example, the fact that a message can go *from A to B* is represented by the ordered pair (A, B), or equivalently, by the line $\overrightarrow{AB}$, as in Fig. 2. Similarly, the fact that A and D choose each other is represented by the two directed lines $\overrightarrow{AD}$ and $\overrightarrow{DA}$.

A *signed graph*, or briefly an *s-graph*, is obtained from a graph when one regards some of the lines as positive and the remaining lines as negative. Considered as a geometric representation of binary relations, an s-graph serves to depict situations or structures in which both a relation and its opposite may occur, e.g., like and dislike. Figure 3 depicts an s-graph, employing the convention that solid lines are positive and dashed lines negative; thus A and B are represented as liking each other while A and C dislike each other.

Combining the concepts of digraph and s-graph, we obtain that of an s-digraph. A *signed digraph*, or an *s-digraph*, is obtained from a digraph by taking some of its lines as positive and the rest as negative.

A *graph of type 2* (**8**), introduced to depict structures in which two differ-

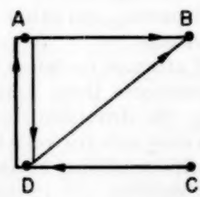


FIG. 2. A directed graph of four points and five directed lines. An $\overrightarrow{AB}$ line indicates the existence of a specified ordered relationship involving the two entities A and B. Thus, for example, if A and B are two people, the $\overrightarrow{AB}$ line might indicate that a message can go from A to B or that A chooses B.

ent relations defined on the same set of elements occur, is obtained from a graph by regarding its lines as being of two different colors (say), and by permitting the same pair of points to be joined by two lines if these lines have different colors. A *graph of type* τ, $\tau = 1, 2, 3, \cdots$, is defined similarly. In an s-graph or s-digraph of type 2, there may occur lines of two different types in which a line of either color may be positive or negative. An example of an s-graph of type 2 might be one depicting for the same *P-O-X* unit both **U** and **L** relations among the entities, where the sign of these relations is indicated.

A *path* is a collection of lines of a graph of the form $AB, BC, \cdots, DE$, where the points $A, B, C, \cdots, D, E$, are distinct. A *cycle* consists of the above path together with the line EA. The *length* of a cycle (or path) is the number of lines in it; an *n-cycle* is a cycle of length n. Analogously to graphs, a *path of a digraph* consists of directed lines of the form $\overrightarrow{AB}, \overrightarrow{BC}, \cdots$, $\overrightarrow{DE}$, where the points are distinct. A *cycle* consists of this path together with the line $\overrightarrow{EA}$. In the later discussion of balance of an s-digraph we shall use the concept of a semicycle. A *semicycle* is a collection of lines obtained by taking exactly one from each pair $\overrightarrow{AB}$ or $\overrightarrow{BA}$, $\overrightarrow{BC}$ or $\overrightarrow{CB}, \cdots$, $\overrightarrow{DE}$ or $\overrightarrow{ED}$, and $\overrightarrow{EA}$ or $\overrightarrow{AE}$. We illustrate semicycles with the digraph of Fig. 2. There are three semicycles in this digraph: $\overrightarrow{AD}, \overrightarrow{DA}$; $\overrightarrow{AD}, \overrightarrow{DB}, \overleftarrow{BA}$; and $\overleftarrow{AD}, \overleftarrow{DB}, \overleftarrow{BA}$. The last two of these semicycles are not cycles. Note that every cycle is a semicycle, and a semicycle of length 2 is necessarily a cycle.

BALANCE

With these concepts of graphs, digraphs, and signed graphs we may now develop a rigorous generalization of Heider's concept of balance.

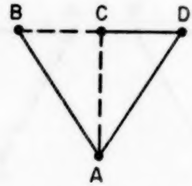


FIG. 3. A signed graph of four points and five lines. Solid lines have a positive sign and dashed lines a negative sign. If the points stand for people and the lines indicate the existence of a liking relationship, this s-graph shows that A and B have a relationship of liking, A and C have one of disliking, and B and D have a relationship of indifference (neither liking or disliking).

It should be evident that Heider's terms, *entity*, *relation*, and *sign of a relation* may be coordinated to the graphic terms, *point*, *directed line*, and *sign of a directed line*. Thus, for example, the assertion that P likes O (P**L**O) may be depicted as a directed line of positive sign $\overrightarrow{PO}$. It should also be clear that Heider's two different kinds of relations (**L** and **U**) may be treated as lines of different type. It follows that a graphic representation of a *P-O-X* unit having positive or negative **L** and **U** relations will be an s-digraph of type 2.

For simplicity of discussion we first consider the situation containing only symmetric relations of a single type (i.e., an s-graph of type 1). Figure 4 shows four such s-graphs. It will be noted that each of these s-graphs contains one cycle: AB, BC, CA. We now need to define the sign of a cycle. The *sign of a cycle* is the product of the signs of its lines. For convenience we denote the sign of a line by $+1$ or -1 when it is positive or negative. With this definition we see that the cycle, AB, BC, CA is positive in s-graph a $(+1 \cdot +1 \cdot +1)$, positive in s-graph b $(+1 \cdot -1 \cdot -1)$, negative

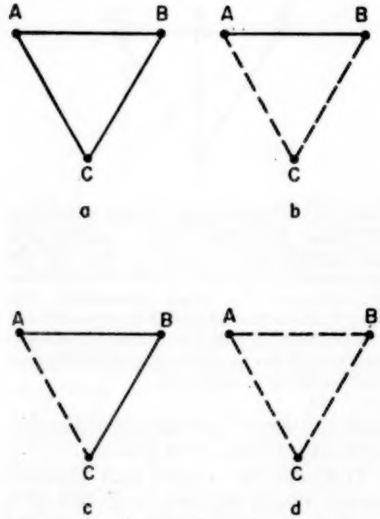


FIG. 4. Four s-graphs of three points and three lines each. Structure *a* and *b* are balanced, but *c* and *d* are not balanced.

in s-graph *c* $(+1 \cdot +1 \cdot -1)$, and negative in s-graph *d* $(-1 \cdot -1 \cdot -1)$. To generalize, a *cycle is positive* if it contains an even number of negative lines, and it is *negative* otherwise. Thus, in particular, a cycle containing only positive lines is positive, since the number of negative lines is zero, an even number.

In discussing the concept of balance, Heider states (see **9**, p. 110) that when there are three entities a balanced state exists if all three relations are positive or if two are negative and one positive. According to this definition, s-graphs *a* and *b* are balanced while s-graphs *c* and *d* are not (Fig. 4). We note that in the examples cited Heider's balanced state is depicted as an s-graph of three points whose cycle is positive.

In generalizing Heider's concept of balance, we propose to employ this characteristic of balanced states as a general criterion for balance of structures with any number of entities. Thus we define an *s-graph* (containing any number of points) as *balanced* if all of its cycles are positive.

Figure 5 illustrates this definition for four s-graphs containing four points. In each of these s-graphs there are seven cycles: *AB*, *BC*, *CA*; *AB*, *BD*, *DA*; *BC*, *CD*, *DB*; *AC*, *CD*, *DA*; *AB*, *BC*, *CD*, *DA*; *AB*, *BD*, *DC*, *CA*; and *BC*, *CA*, *AD*, *DB*. It will be seen that in s-graphs *a* and *b* all seven cycles are positive, and these s-graphs are therefore balanced. In s-graphs *c* and *d* the cycle, *AB*, *BC*, *CA*, is negative (as are several others), and these s-graphs are therefore not balanced. It is obvious that this definition of balance is applicable to structures containing any number of entities.[8]

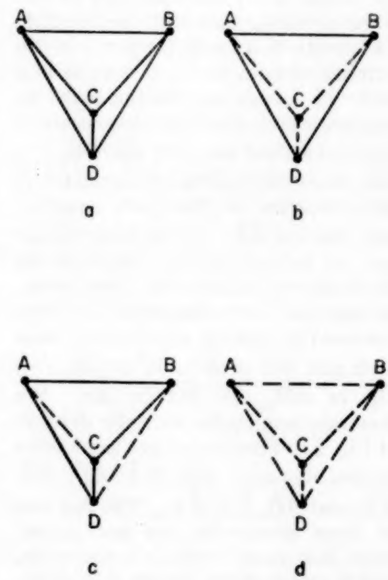


FIG. 5. Four s-graphs containing four points and six lines each. Structures *a* and *b* are balanced, but *c* and *d* are not balanced.

[8] If an s-graph contains no cycles, we say that it is "vacuously" balanced, since all (in this case, none) of its cycles are positive.

The extension of this definition of balance to s-digraphs containing any number of points is straightforward. Employing the same definition of *sign of a semicycle* for an s-digraph as for an ordinary s-graph, we similarly define an s-digraph as balanced if all of its semicycles are positive.

Consider now Heider's *P-O-X* unit, containing two persons P and O and an impersonal entity X, in which we are concerned only with liking relations. Figure 6 shows three of the possible 3-point s-digraphs which may represent such *P-O-X* units. A positive $\overrightarrow{PO}$ line means that P likes O, a negative $\overrightarrow{PO}$ line means that P dislikes O. We assume that a person can like or dislike an impersonal entity but that an impersonal entity can neither like nor dislike a person.[9] We also rule out of consideration here "ambivalence," where a person may simultaneously like and dislike another person or impersonal entity.

In each of these s-digraphs there are three semicycles: $\overrightarrow{PO}$, $\overrightarrow{OP}$; $\overrightarrow{PO}$, $\overrightarrow{OX}$, $\overleftarrow{XP}$; and $\overleftarrow{PO}$, $\overrightarrow{OX}$, $\overleftarrow{XP}$. If we confine our discussion to the kind of structures represented in Fig. 6 (i.e., where there is no ambivalence and where all possible positive or negative lines are present), it will be apparent that: when P and O like each other, the s-digraph is balanced only if both persons either like or dislike X (s-digraph a is not balanced); when P and O dislike each other, the s-digraph is balanced only if one person likes X and the other person dislikes X (s-digraph b is balanced); and when one person likes the other but the other dislikes him, the s-digraph must be not balanced (s-digraph c is not balanced). These conclusions are consistent with Heider's discussion of *P-O-X* units and with Newcomb's treatment of the *A-B-X* model.

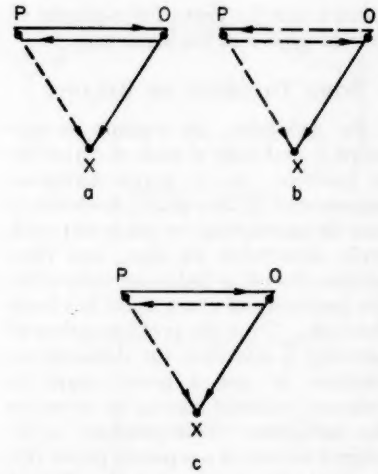


FIG. 6. Three s-digraphs representing Heider's *P-O-X* units. Only structure b is balanced.

The further extension of the notion of balance to s-graphs of type 2 remains to be made. The simplest procedure would be simply to ignore the types of lines involved. Then we would again define an *s-graph of type 2* to be *balanced* if all of its cycles are positive. This definition appears to be consistent with Heider's intention, at least as it applies to a situation containing only two entities. For in speaking of such situations having both **L** and **U** relations, he calls them balanced if both relations between the same pair of entities are of the same sign (see **9**, p. 110). There remains some question as to whether this definition will fit empirical findings for cycles of greater length. Until further evidence is available, we advance the above formulation as a tentative definition. Obviously the definition of

[9] In terms of digraph theory we define an *object* as a *point with zero output*. Thus a completely indifferent person is an object. If, psychologically, an impersonal entity is active and likes or dislikes a person or another impersonal entity, then in terms of digraph theory it is not an object.

balance can be given for s-graphs of general type τ in the same way.

SOME THEOREMS ON BALANCE

By definition, an s-graph is balanced if and only if each of its cycles is positive. In a given situation represented by an s-graph, however, it may be impractical to single out each cycle, determine its sign, and then declare that it is balanced only after the positivity of every cycle has been checked. Thus the problem arises of deriving a criterion for determining whether or not a given graph is balanced without having to revert to the definition. This problem is the subject matter of a separate paper (**6**), in which two necessary and sufficient conditions for an s-graph to be balanced are developed. The first of these is no more useful than the definition in determining by inspection whether an s-graph is balanced, but it does give further insight into the notion of balance. Since the proofs of these theorems may be found in the other paper, we shall not repeat them here.

Theorem. An s-graph is balanced if and only if all paths joining the same pair of points have the same sign.

Thus, we can ascertain that the s-graph of Fig. 7 is balanced either by listing each cycle separately and verifying that it is positive, or, using this

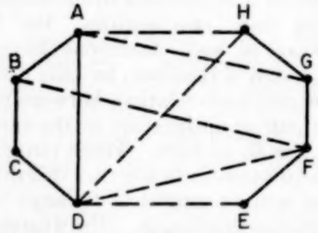


FIG. 7. An s-graph of eight points and thirteen lines which, by aid of the structure theorem, can be readily seen as balanced.

theorem, by considering each pair of possible points and verifying that all possible paths joining them have the same sign. For example, all the paths between points A and E are negative, all paths joining A and C are positive, etc.

The following structure theorem has the advantage that it is useful in determining whether or not a given s-graph is balanced without an exhaustive check of the sign of every cycle, or of the signs of all paths joining every pair of points.

Structure theorem. An s-graph is balanced if and only if its points can be separated into two mutually exclusive subsets such that each positive line joins two points of the same subset and each negative line joins points from different subsets.

Using the structure theorem, one can see at a glance that the s-graph of Fig. 7 is balanced, for A, B, C, D, and E, F, G, H are clearly two disjoint subsets of the set of all points which satisfy the conditions of the structure theorem.

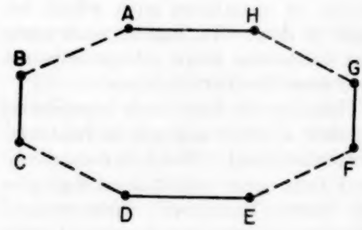
It is not always quite so easy to determine balance of an s-graph by inspection, for it is not always necessarily true that the points of each of the two subsets are connected to each other. Thus the two s-graphs of Fig. 8 are balanced, even though neither of the two disjoint subsets is a connected subgraph. However, the structure theorem still applies to both of the s-graphs of Fig. 8. In the first graph the appropriate subsets of points are A, D, E, H and B, C, F, G; while in the second one we take A_1, B_1, A_3, B_3, A_5, B_5 and A_2, B_2, A_4, B_4.

In addition to providing two necessary and sufficient conditions for balance, these theorems give us further information about the nature of balance. Thus if we regard the s-graph as representing Heider's **L**-relation in

a group, then the structure theorem tells us that the group is necessarily decomposed into two subgroups (cliques) within which the relationships that occur are positive and between which they are negative. The structure theorem, however, does not preclude the possibility that one of the two subsets may be empty—as, for example, when a connected graph contains only positive lines.

The first theorem also leads to some interesting consequences. Suppose it were true, for example, that when two people like each other they can influence each other positively (i.e., produce intended changes in the other), but when two people dislike each other they can only influence each other negatively (i.e., produce changes opposite to those intended). An s-graph depicting the liking relations among a group of people will, then, also depict the potential influence structure of the group. Suppose that Fig. 7 represents such a group. If *A* attempts to get *H* to approve of something, *H* will react by disapproving. If *H* attempts, in turn, to get *G* to disapprove of the same thing, he will succeed. Thus *A*'s (indirect) influence upon *G* is negative. The first theorem tells us that *A*'s influence upon *G* must be negative, regardless of the path along which the influence passes, since the s-graph is balanced. In general, the sign of the influence exerted by any point upon any other will be the same, no matter what path is followed, since the graph is balanced.

By use of the structure theorem it can be shown that in a balanced group any influence from one point to another within the same clique must be positive, even if it passes through individuals outside of the clique, and the influence must be negative if it goes from a person in one clique to a person in the other. (It should be noted that in this discussion we give the term "clique" a special meaning, as above.) Thus, under the assumed conditions, any exerted influence regarding opinions will tend to produce homogeneity within cliques and opposing opinions between cliques.



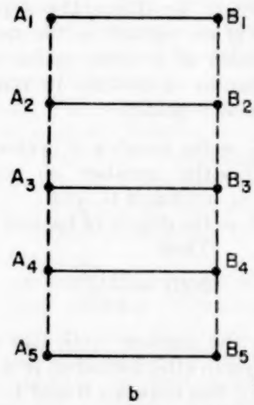


FIG. 8. Two s-graphs whose balance cannot be determined easily by visual inspection.

Although we have illustrated these theorems by reference to social groups, it should be obvious that they hold for any empirical realizations of s-graphs.

FURTHER CONCEPTS IN THE THEORY OF BALANCE

The concepts of balance as developed up to this point are clearly oversimplifications of the full com-

plexity of situations with which we want to deal. To handle such complex situations more adequately, we need some further concepts.

Thus far we have only considered whether a given s-graph is balanced or not balanced. But it is intuitively clear that some unbalanced s-graphs are "more balanced" than others! This suggests the introduction of some scale of balance, along which the "amount" of balance possessed by an unbalanced s-graph may be measured. Accordingly we define the ***degree of balance of an s-graph*** as the ratio of the number of positive cycles to the total number of cycles. In symbols, let G be an s-graph,

$c(G)$ = the number of cycles of G,
$c_+(G)$ = the number of positive cycles of G, and
$b(G)$ = the degree of balance of G.
Then

$$b(G) = \frac{c_+(G)}{c(G)}.$$

Since the number $c_+(G)$ can range from zero to $c(G)$ inclusive, it is clear that $b(G)$ lies between 0 and 1. Obviously $b(G) = 1$ if and only if G is balanced. We can give the number $b(G)$ the following probabilistic interpretation: the degree of balance of an s-graph is the probability that a randomly chosen cycle is positive.

Does $b(G) = 50\%$ mean that G is exactly one-half balanced? The

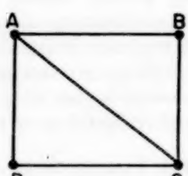


FIG. 9. A graph of four points which can acquire degrees of balance of only .33 and 1.00 regardless of the assignment of positive and negative signs to its five lines.

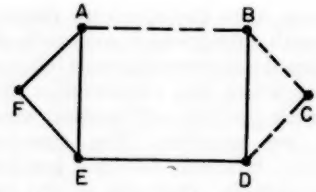


FIG. 10. An s-graph which is 3-balanced but not 4-balanced.

answer to this question depends on the possible values which $b(G)$ may assume. This in turn depends on the structure of the s-graph G. Thus, if G is the complete graph of 3 points and G is not balanced, then the only possible value is $b(G) = 0$, since there is only one cycle. Similarly, if the lines of G are as in Fig. 9, some of which may be negative, and if G is not balanced, then the only possible value is $b(G) = \frac{1}{3}$, and $b(G) = 50\%$ does not even occur for this structure. Thus any interpretation given to the numerical value of $b(G)$ must take into account the distribution curve for $b(G)$, which is determined by the structure of G.

We now consider the corresponding concept of the degree of balance for s-diagraphs. Since an s-digraph is balanced if all of its semicycles are positive, the ***degree of balance of an s-digraph*** is taken as the ratio of the number of positive semicycles to the total number of semicycles.

In a given s-graph which represents the signed structure of some psychological situation, it may happen that only cycles of length 3 and 4 are important for the purpose of determining balance. Thus in an s-graph representing the relation **L** in a complex group, it will not matter at all to the group as a whole whether a cycle of length 100, say, is positive. To handle this situation rigorously, we define an s-graph to be ***N-balanced*** if all its cycles of length not exceeding

N are positive. Of course the degree of N-balance is definable and computable for any s-graph. Examples can be given of unbalanced s-graphs which are, however, N-balanced for some N. Figure 10 illustrates this phenomenon for $N = 3$, since all of its 3-cycles are positive, but it has a negative 4-cycle.[10]

For certain problems, one may wish to concentrate only on one distinguished point and determine whether an s-graph is balanced there. This can be accomplished by the notion of local balance. We say an s-graph is *locally balanced at point P* if all cycles through P are positive. Thus the s-graph of Fig. 11 is balanced at points A, B, C, and not balanced at D, E, F. If this figure represents a sociometric structure, then the concept of local balance at A is applicable provided A is completely unconcerned about the relations among D, E, F.

Some combinatorial problems suggested by the notions of local balance and N-balance have been investigated by Harary (**7**). The principal theorem on local balance, which follows, uses the term "articulation point" which we now define. An *articulation point*[11] of a connected graph is one whose removal[12] results in a disconnected graph. Thus the point D is the only articulation point of Fig. 11. We now state the main theorem on local balance, without proof.

Theorem. If a connected s-graph G is balanced at P, and Q is a point on a cycle passing through P, where Q is not an articulation point, then G is also balanced at Q.

Figure 11 serves to illustrate this theorem, for the s-graph is balanced at A, and is also balanced at B but is not balanced at D, which is an articulation point.

In actual practice, both local balance and N-balance may be employed. This can be handled by introducing the combined concept of local N-balance. Formally we say that an s-graph is *locally N-balanced at P* if all cycles of length not exceeding N and passing through P are positive. Obviously the degree of local N-balance can be defined analogously to the degree of balance.

In summary, the concept of degree of balance removes the limitation of dealing with only balanced or unbalanced structures, and in addition is susceptible to probabilistic and statistical treatment. The definition of local balance enables one to focus at any particular point of the structure. The introduction of N-balance frees us from the necessity of treating all cycles as equally important in determining structural balance. Thus, the extensions of the notion of balance

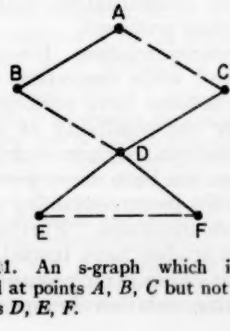


FIG. 11. An s-graph which is locally balanced at points A, B, C but not balanced at points D, E, F.

[10] One way of viewing the definition of N-balance is to regard cycles of length N as having weight 1, and all longer cycles as of weight 0. Of course, it is possible to generalize this idea by assigning weights to each length, e.g., weight $1/2^n$ to length N.

[11] A characterization of the articulation points of a graph, or in other words the liaison persons in a group, is given by Ross and Harary (**19**), using the "structure matrix" of the graph. An exposition of this concept is given in Harary and Norman (**8**).

[12] By the removal of a point of a graph is meant the deletion of the point and all lines to which it is incident.

developed in this section permit a study of more complicated situations than does the original definition of Heider.

ADEQUACY OF THE GENERAL THEORY OF BALANCE

In any empirical science the evaluation of a formal model must be concerned with both its formal properties and its applicability to empirical data. An adequate model should account for known findings in a rigorous fashion and lead to new research. Although it is not our purpose in this article to present new data concerning tendencies toward balance in empirical systems, we may attempt to evaluate the adequacy of the proposed general theory of balance in the light of presently available research.

Our review of Heider's theory of balance and of the research findings related to it has revealed certain ambiguities and limitations concerning (*a*) the treatment of unsymmetric relations; (*b*) the generalization to systems containing more than three entities; (*c*) the distinction between the complement and the opposite of a relation; (*d*) the simultaneous existence of relationships of different types; and (*e*) the applicability of the concept of balance to empirical systems other than cognitive ones. We now comment briefly upon the way in which our generalization deals with each of these problems.

Unsymmetric relations. It was noted above that, while theoretical discussions of balance have sometimes allowed for the possibility of unsymmetric relations, no rigorous definition of balance has been developed to encompass situations containing unsymmetric relationships. Furthermore, empirical studies have tended to assume that liking is reciprocated, that each liking relation is symmetric. By stating the definition of balance in terms of s-digraphs, we are able to include in one conceptual scheme both symmetric and unsymmetric relationships. And it is interesting to observe that, according to this definition, whenever the lines $\overrightarrow{PO}$ and $\overrightarrow{OP}$ are of different signs, the s-digraph containing them is not balanced. Thus, to the extent that tendencies toward balance have been effective in the settings empirically studied, the assumption of symmetry has, in fact, been justified.

Situations containing any finite number of entities. Heider's discussion of balance has been confined to structures containing no more than three entities. The definition of balance advanced here contains no such limitation; it is applicable to structures containing any finite number of entities. Whether or not empirical theories of balance will be confirmed by research dealing with larger structures can only be determined by empirical work. It is clear, however, that our generalization is consistent with the more limited definition of Heider.

A relation, its complement, and its opposite. Using s-graphs and s-digraphs to depict relationships between entities allows us to distinguish among three situations: the presence of a relation (positive line), the presence of the opposite of a relation (negative line), and the absence of both (no line). The empirical utilization of this theory requires the ability to distinguish among these three situations. In our earlier discussion of the literature on balance, we noted, however, a tendency to distinguish only the presence or absence of a relationship. It is not always clear, therefore, in attempting to depict previous research in terms of s-graph theory whether a given empirical relationship

should be coordinated to no line or to a negative line.

The experiment of Jordan (11) illustrates this problem quite clearly. He employed three entities and specified certain **U** and **L** relations between each pair of entities. The empirical realization of these relations was obtained in the following way: **U** was made into "has some sort of bond or relationship with"; $\sim$**U** into "has no sort of bond or relationship with;" **L** was made into "like;" and $\sim$**L** into "dislike." Viewed in the light of s-graph theory, it would appear that Jordan created s-graphs of type 2 (which may contain positive and negative lines of type **U** and type **L**). It would also appear, however, that the $\sim$**U** relation should be depicted as the absence of any **U**-line but that the $\sim$**L** relation should be depicted as a negative **L**-line. If this interpretation is correct, Jordan's classification of his situations as "balanced" and "imbalanced" will have to be revised. Instead of interpreting the $\sim$**U** relation as a negative line, we shall have to view it as no **U**-line, with the result that all of his situations containing $\sim$**U** relations are vacuously balanced by our definition since there are no cycles.

It is interesting to examine Jordan's data in the light of this reinterpretation. He presented subjects with 64 hypothetical situations, half of which were "balanced" and half "imbalanced" by his definition. He had subjects rate the degree of pleasantness or unpleasantness experienced in each situation (a high score indicating unpleasantness). For "balanced" situations the mean rating was 46 and for "imbalanced" ones, 57.

If, however, we interpret Jordan's $\sim$**U** relation as the absence of a line, his situations must be reclassified. Of his 32 "balanced" situations, 14 have no $\sim$**U** relation and thus remain balanced. The mean unpleasantness score for these is 39. The remaining 18 of his "balanced" situations, having at least one $\sim$**U** relation, become vacuously balanced since no cycle remains. The mean unpleasantness of these vacuously balanced situations is 51. Of Jordan's 32 "imbalanced" situations, 19 contain at least one $\sim$**U** relation, thus also becoming vacuously balanced, and the mean unpleasantness score for these is 51. The remaining 13 situations, by having no $\sim$**U** relations, remain imbalanced, and their mean score is 66. Thus it is clear that the difference in pleasantness between situations classed by Jordan as "balanced" and "imbalanced" is greatly increased if the vacuously balanced situations are removed from both classes (balanced, 39; vacuously balanced, 51; not balanced, 66). These findings lend support to our view that the statement "has no sort of bond or relationship with" should be represented as the absence of a line.[13]

Relations of different types. A basic feature of Heider's theory of balance is the designation of two types of relations (**L** and **U**). Our generalization of the definition of balance permits the inclusion of any number of types of relations. Heider discusses the combination of types of relations only for the situation involving two entities, and it is clear that our definition is consistent with his within this limitation. It is interesting to note

[13] A strict test of our interpretation of Jordan's data is not possible since he specified for any given pair of entities only either the **L** or **U** relation. We can but guess how the subjects filled in the missing relationship. In the light of our discussion of relations of different types, in the next section, it appears that subjects probably assumed a positive unit relation when none was specified, since there is a marked tendency to experience negative liking relations as unpleasant.

that Jordan (**11**) finds positive liking relations to be experienced as more pleasant than negative ones. This finding may be interpreted as indicating a tendency toward "positivity" over and above the tendency toward balance. It is possible, however, that in the hypothetical situations employed by Jordan the subjects assumed positive unit relations between each pair of entities. If this were in fact true, then a positive liking relation would form a positive cycle of length 2 with the positive unit relation, and a negative liking relation would form a negative cycle of length 2 with the positive unit relation. And, according to the theory of balance, the positive cycle should produce more pleasantness than the negative one. This interpretation can be tested only through further research in which the two relations are independently varied.

Empirical applicability of concept of balance. Heider's discussion of balance refers to a cognitive structure, or the life space of a single person. Newcomb suggests that a similar conception may be applicable to interpersonal systems objectively described. Clearly, our definition of balance may be employed whenever the terms "point" and "signed line" can be meaningfully coordinated to empirical data of any sort. Thus, one should be able to characterize a communication network or a power structure as balanced or not. Perhaps it would be feasible to use the same definition in describing neural networks. It must be noted, however, that it is a matter for empirical determination whether or not a tendency to achieve balance will actually be observed in any particular kind of situation, and what the empirical consequences of not balanced configurations are. Before extensive utilization of these notions can be accomplished, certain further conceptual problems regarding balance must be solved.

One of the principal unsolved problems is the development of a systematic treatment of relations of varying strength. We believe that it is possible to deal with the strength of relations by the concept of a graph of strength σ, suggested by Harary and Norman (**8**).

SUMMARY

In this article we have developed a generalization of Heider's theory of balance by use of concepts from the mathematical theory of linear graphs. By defining balance in graph-theoretic terms, we have been able to remove some of the ambiguities found in previous discussions of balance, and to make the concept applicable to a wider range of empirical situations than was previously possible. By introducing the concept *degree of balance*, we have made it possible to treat problems of balance in statistical and probabilistic terms. It should be easier, therefore, to make empirical tests of hypotheses concerning balance.

Although Heider's theory was originally intended to refer only to cognitive structures of an individual person, we propose that the definition of balance may be used generally in describing configurations of many different sorts, such as communication networks, power systems, sociometric structures, systems of orientations, or perhaps neural networks. Only future research can determine whether theories of balance can be established for all of these configurations. The definitions developed here do, in any case, give a rigorous method for describing certain structural aspects of empirical configurations.

REFERENCES

1. BACK, K. The exertion of influence through social communication. *J. abnorm. soc. Psychol.*, 1951, **46**, 9–23.
2. BAVELAS, A. A mathematical model for group structures. *Appl. Anthrop.*, 1948, **7**, No. 3, 16–30.
3. FESTINGER, L. The analysis of sociograms using matrix algebra. *Hum. Relat.*, 1949, **2**, 153–158.
4. FESTINGER, L., SCHACHTER, S., & BACK, K. *Social pressures in informal groups.* New York: Harper, 1950.
5. FESTINGER, L., & THIBAUT, J. Interpersonal communication in small groups. *J. abnorm. soc. Psychol.*, 1951, **46**, 92–99.
6. HARARY, F. On the notion of balance of a signed graph. *Mich. math. J.*, 1953–54, **2**, 143–146.
7. HARARY, F. On local balance and N-balance in signed graphs. *Mich. math. J.*, 1955–56, **3**, 37–41.
8. HARARY, F., & NORMAN, R. Z. *Graph theory as a mathematical model in social science.* Ann Arbor, Mich.: Institute for Social Research, 1953.
9. HEIDER, F. Attitudes and cognitive organization. *J. Psychol.*, 1946, **21**, 107–112.
10. HOROWITZ, M. W., LYONS, J., & PERLMUTTER, H. V. Induction of forces in discussion groups. *Hum. Relat.*, 1951, **4**, 57–76.
11. JORDAN, N. Behavioral forces that are a function of attitudes and of cognitive organization. *Hum. Relat.*, 1953, **6**, 273–287.
12. KOFFKA, K. *Principles of gestalt psychology.* New York: Harcourt, Brace, 1935.
13. KÖHLER, W. *Dynamics in psychology.* New York: Liveright, 1940.
14. KÖNIG, D. *Theorie der endlichen und unendlichen Graphen.* New York: Chelsea, 1950 (originally published in Leipzig, 1936).
15. LEWIN, K. *Field theory in social science.* New York: Harper, 1951.
16. NEWCOMB, T. M. *Personality and social change.* New York: Dryden, 1943.
17. NEWCOMB, T. M. An approach to the study of communicative acts. *Psychol. Rev.*, 1953, **60**, 393–404.
18. PARSONS, T. *Essays in sociological theory.* (Rev. Ed.) Glencoe, Ill.: Free Press, 1954.
19. ROSS, I. C., & HARARY, F. Identification of the liaison persons of an organization using the structure matrix. *Mgmt Sci.*, 1955, **1**, 251–258.
20. SCHACHTER, S. Deviation, rejection and communication. *J. abnorm. soc. Psychol.*, 1951, **46**, 190–207.
21. TAGIURI, R., BLAKE, R. R., & BRUNER, J. S. Some determinants of the perception of positive and negative feelings in others. *J. abnorm. soc. Psychol.*, 1953, **48**, 585–592.

(Received January 11, 1956)

Psychological Review
Vol. 63, No. 5, 1956

RORSCHACH'S AFFECT–COLOR HYPOTHESIS AND ADAPTATION–LEVEL THEORY

CLAY E. GEORGE AND WARREN C. BONNEY

Agricultural and Mechanical College of Texas

The intent of the present effort is to re-evaluate some concepts of emotional reactivity in the light of Helson's adaptation-level (AL) theory (6). Any inclusive behavior theory should be applicable to emotional behavior, as well as to other forms of behavior; more specifically, it should also be applicable to such clinically observed behavior as Rorschach color responses. Rorschach color responses are presumed to be perceptual correlates of emotional reactivity. We are, in effect, attempting to order an array of clinical facts into an existing, systematic theory of behavior. By so doing, the application of AL theory may be broadened to include a new body of data.

Rorschach responses, from which overt behavior is predicted, are characteristically seen as being symbolic of unconscious needs, and the processes mediating between the latter and behavior must somehow be accounted for. AL theory is particularly concerned with the explanation of intervening variables. As such, we believe that it offers the most promising framework in which to investigate the data of emotional reactivity.

In order to investigate the adequacy of AL theory for this purpose, we will attempt to explain some representative findings concerning Rorschach color data. Response to color was chosen since color is a relatively uncomplicated stimulus which yields a response with affective meaning. Rorschach color responses provide a considerable body of data upon which one may draw.

Rorschach, in his report on the *Psychodiagnostic,* formulated what has come to be known as the "affect-color" theory. As Rorschach expressed it,

"It has long been realized that there must exist a very close relationship between color and affectivity. The gloomy person is one to whom everything looks 'black,' while the cheerful person is said to see everything through rose-colored glasses" (**14**, p. 99).

Beck (**1**) seems to have largely accepted Rorschach's formulation. Klopfer (**8, 9**) has also accepted the essence of Rorschach's affect-color theory but has considerably elaborated upon it. Fortier (**2**) recently reviewed the literature and concluded that the acceptance of a roughly one-to-one relationship between color and affect is defensible.

Perhaps the most damaging evidence against a simple affect-color theory is provided by the experiment of Siipola (**15**). Her major finding was that it is the perception of color inappropriate to the form enclosing that color which gives rise to emotional response, not color as such. McClelland (**11**) offers a corollary to AL theory which seems to be particularly relevant to Siipola's findings. To paraphrase McClelland: People are indifferent to (bored by) that which they wholly expect, find pleasing that which is somewhat unexpected (novel), and find unpleasant that which radically differs from expectation (shocking). Siipola's findings are easily handled in these terms. When the subject (*S*) perceives a form representing, to him, a human face, he experiences "something radically different from expectation" when that face is bright green in color. The shock may be similar to that which one might undergo upon be-

ing confronted by the notorious purple cow.

RELEVANT EXPERIMENTAL FINDINGS

A sample of college students has been reported by one of the writers (5) to have ranked the colored cards in the order of preference as follows: Cards VIII and X tied for the first rank (best liked), then Cards III and IX, and finally II. The achromatic cards ranked below the colored. In terms of McClelland's formulation we would say that the colored cards were novel, and therefore pleasant to the *S*s. Card II would be unpleasing to *S*s with low ALs since it is the first colored card, and it comes after Card I has set the *S* to expect achromatic stimuli. Card II, in its standard colored form, was reported to be relatively unpleasant by a group of unstable subjects studied by Wallen (18).

Hershenson (7) has found that adolescent females find Card II less pleasing than do males of the same age. She implies that this may be due to a sort of "menstrual anxiety" on the part of these adolescent females. Aside from this, her *S*s reacted in a manner very similar to that of George's college students. To elicit shock reactions from ink-blot stimuli, Siipola (16) had first to lower her subjects' ALs by subjecting them to a press situation. She aroused a degree of anxiety in the *S*s and they reacted with color shock, a reaction similar to that found in neurotic *S*s who had chronically low ALs (12, 18). Finally, Mitchell (12) has reported that the majority of psychoneurotic *S*s ranking Card II as the least liked of the cards state that they did so because the color was unpleasant to them.

In terms of AL theory, *S*s with presumably normally high ALs, George's college students, Hershenson's high school males, and Wallen's stable males —tend to report the unexpected emergence of color on Card II as pleasant. On the other hand, the unexpected color is reported as unpleasant to *S*s with either chronically low or experimentally lowered ALs. Such *S*s include Wallen's unstable males, Mitchell's psychoneurotics, Hershenson's adolescent females, and Siipola's college students under press conditions. Other pertinent data are provided by Holzberg and Wexler (8), who found that reaction time to colored cards is significantly lower in schizophrenic *S*s upon retest than it was on the first test. We might explain this finding by saying that the *S*s were expecting colored stimulation on the retest and were therefore less surprised or shocked by it. Likewise, we could explain the marked tendency of normal *S*s to prefer Cards VIII, IX and X (5) by assuming that Cards II and III had prepared them for colored stimuli, thereby offsetting possible surprise reactions, but that they were still pleasantly novel, since Cards IV and VII preceded them.

From the foregoing discussion we would be led to expect different affective reactions to the cards when they are presented in reverse order. Data illustrating the fact that different reactions are thus elicited are provided by a further phase of Wallen's experiment. One group of stable men were presented the cards in standard order, another group in reverse order (Cards X through I). The percentage reported by Wallen as liking each card were, for standard and reverse order groups respectively: Card I—15, 43; II—23, 37; III—23, 35; IV—34, 23; V—58, 52; VI—49, 54; VII—26, 14; VIII—72, 54; IX—62, 38; X—69, 23. For the differences between group reactions to Cards I, VIII, IX and X, Wallen reports CRs of 3.7, 2.2, 2.8, and 6.8, respectively. Assuming that *S*s come to the Rorschach cards with an expectation of achromatic

stimuli, as persons in our culture expect black and white movies when color is not advertised, we can explain the significant difference between the groups in their affective reactions to Card X.

In the standard order, Card X is well liked because it is preceded by four colored cards and the *S*s are prepared for further chromatic stimuli. In the reverse order, Card X is less well liked because *S*s are not expecting color, and are surprised or shocked by it. The reverse-order group reported less positive affect for Cards IX and VIII than did the standard group, but they did report an increasing liking for chromatic cards as they became adapted to color stimulation. Furthermore, the reverse group liked Card II better than did the standard group, and they liked Card I significantly more, since it came after the colored Cards III and II and was therefore novel.

To extend further the consideration of color phenomena we may appeal to Helson's concepts of residual and pooling. In this way we may attempt to explain those cases in which *S*s show persistent shock reaction to color even after there has been ample opportunity to adapt, i.e., even though the experience of color is no longer unexpected.

According to Helson (6, p. 382) there are three major factors in adaptation, of which two are relevant to the present discussion: the stimulus attended to or being judged (the testing situation and any preceding plates), and effects of previous stimuli forming the residual stimulus (whatever latent reaction processes toward color the subject may bring to the testing situation). Thus, persisting tendencies toward color-shock reactions to the last three colored cards, or even to retest situations, can be easily fitted into AL theory.

Next, we must explain the data derived from work with the formal scoring categories utilizing color. We may speculate that Form-Color (*FC*) responses occur only when the *S* does not react with displeasure, or shock, to the stimulus, and usually only when he reacts with pleasure. In terms of AL theory any emotionality arising from unexpected, or residual, factors associated with the stimulus is sufficiently close to the *S*'s zero point of adaptation to allow him to maintain intellectual efficiency, and he is thus able to order the stimuli in the more easily structured form dominant combinations.

The Color-Form (*CF*) responder is so influenced by his emotional response to unexpected or residual associations with color that he is unable to make full use of his intellectual capacity to order properly the form properties of the blot.

Finally, the Color (*C*) responder is overwhelmed by the unexpected and residual associations, and is momentarily deprived of any capacity for structuring the form elements of the stimulus. It would be difficult to conceive of a *C* response in which residual components are not of primary importance. Such responses are not expected in adult *S*s unless the residuals are such that AL is extremely low for color, or for specific associations with the color perceived, e.g., blood responses. When the *S* is totally overwhelmed, he simply rejects the card.

Helson's concept of stimulus pooling may have an interesting application to startle or shock reaction. Helson cites experimental evidence (6, p. 381) to confirm his position that stimulus pooling may occur below the level of conscious awareness. This process of pooling as conceived by Helson would seem to be incomplete for the present effort. One might also appeal to the Freudian concept of repression. Pooling would then for us consist of subliminal effects plus repressed affect. The sudden presentation of a stimulus for which the

S has acquired below-conscious, affect-laden residuals brings about the sudden emergency of pure affect, the nature of which the *S* cannot readily order cognitively. This surge of affect, which is unexpected and inexplicable to the *S*, causes the response which Rorschach referred to as color shock. It is not the surge of affect, per se, but the unexpectedness and inexplicability of it *for the S* that brings about the startle response.

In the light of the above theoretical discussion of the color score categories, how might we explain the findings of Smith (**17**)? This investigator reports low but significant positive correlations between *FC* and Wechsler-Bellevue subtest scores, low but significant negative correlations between *CF* and Wechsler-Bellevue subtest scores. The *FC* responder adapts to any stressful event (test situation) and therefore is able to deal with such events without undue emotional interference. The *CF* responder, on the other hand, with his low AL for any emotional or stressful stimulation, is less able to utilize any intellectual capacity he may possess. This *S* is just as incapable of structuring the components of the test situation as of structuring the form elements of an ink blot. It seems, then, that we are not dealing with intellectual variables so much as ability to cope successfully with stress in test situations. It may be that affective residuals which become associated with the stressful aspects of the test situation are largely responsible for the relative inability of the *CF* responder to make adequate use of his cognitive capacities.

The reported negative relationships of the weighted color score (Sum $C = .5FC + CF + 1.5C$) with various measures of intelligence (**3**, **13**) may be explained in the same way. These relationships may be expected to hold because of the heavy weighting of *CF* and *C;* they might be even higher were not *FC* weighted positively in the formula.

Evidence that *FC Ss* react in an emotionally stable manner, but with some suggestion of stereotypy and inhibition, is found in a report by George (4). This report also indicates that *CF Ss* tend to react with impulsivity. We would say then, that the *CF S*'s affective residuals are of sufficient force to lower his AL for any emotional situation to such an extent that he is able to react with only a minimum of control.

SUMMARY

We have attempted to expand adaptation-level theory to include a body of psychological fact hitherto not subsumed under any systematic theory of adjustive behavior. AL theory has traditionally been applied to psychophysical methods of studying perception. Little or no work has been done on perceptual phenomena associated with affectivity. In our efforts to bring AL theory to bear on such phenomena we have utilized Rorschach color data. It has been possible to explain many findings concerning color responses by reference to AL theory. These findings have not previously been satisfactorily explained by any inclusive theory. In this paper some apparently contradictory findings have been included within a single theoretical system.

REFERENCES

1. BECK, S. J. *Rorschach's test.* New York: Grune & Stratton, 1949, Vol. 1.
2. FORTIER, R. H. The response to color and the ego functions. *Psychol. Bull.*, 1953, **50**, 41–63.
3. GARRISON, M. Relationships between Rorschach scores and clinical changes in mental patients. *J. Pers.*, 1948, **17**, 146–152.
4. GEORGE, C. E. Rorschach color and form variables: a validation study. Unpublished master's thesis, Univer. of Arizona, 1953.

5. GEORGE, C. E. Stimulus value of the Rorschach cards: a composite study. *J. proj. Tech.*, 1955, **19**, 17–20.
6. HELSON, H. (Ed.) *Theoretical foundations of psychology.* New York: Van Nostrand, 1951.
7. HERSHENSON, JEANNE R. Preference of adolescents for Rorschach figures. *Child Develpm.*, 1949, **20**, 101–118.
8. HOLZBERG, J. D., & WEXLER, M. The predictability of schizophrenic performance on the Rorschach test. *J. consult. Psychol.*, 1950, **14**, 395–399.
9. KLOPFER, B., & KELLEY, D. M. *The Rorschach technique.* Yonkers, N. Y.: World Book, 1942.
10. KLOPFER, B., AINSWORTH, M. D., KLOPFER, W. G., & HOLT, R. R. *Developments in the Rorschach technique.* Yonkers, N. Y.: World Book, 1954, Vol. I.
11. McCLELLAND, D. C., ATKINSON, J. W., CLARK, R. A., & LOWELL, E. L. *The achievement motive.* New York: Appleton-Century-Crofts, 1953.
12. MITCHELL, MILDRED B. Preferences for Rorschach cards. *J. proj. Tech.*, 1951, **16**, 203–211.
13. ROE, ANNE. Psychological tests of research scientists. *J. consult. Psychol.*, 1951, **15**, 492–495.
14. RORSCHACH, H. *Psychodiagnostic* (4th Ed.). Berne: Hans Huber, 1949.
15. SIIPOLA, ELSA M. The influence of color on reactions to ink blots. *J. Pers.*, 1950, **18**, 358–382.
16. SIIPOLA, ELSA M., & TAYLOR, VIVIAN. Reactions to ink blots under free and pressure conditions. *J. Pers.*, 1952, **21**, 22–47.
17. SMITH, S. R. The Rorschach examination and general intelligence: a validation study. Unpublished doctor's dissertation, Univer. of Calif., 1951.
18. WALLEN, R. The nature of color shock. *J. abnorm. soc. Psychol.*, 1948, **43**, 346–356.

(Received December 6, 1955)

Psychological Review
Vol. 63, No. 5, 1956

ON ACTIVITY IN THE GOAL REGION [1]

MARY HENLE

Graduate Faculty of Political and Social Science, New School for Social Research

It is the purpose of this paper to call attention to a problem neglected by motivational theories, that of activity in the goal region. Contemporary motivational theories are theories of striving to reach goals; their interest stops once the goal has been achieved. Even hedonism seems to deal only with the pursuit of happiness and to have little place for happiness itself.

Common experience suggests, however, that the task of a theory of motivation does not end with the attainment of the goal. Human beings are not concerned exclusively with trying to reach what they do not have. We may enjoy a friend, for example, without other relevant motivation, without trying to accomplish anything. An activity shared with the loved person is more enjoyable than the same activity carried out with an indifferent partner. A friendly conversation may proceed without ulterior motive. If we read and enjoy a good book, we cannot adequately describe our motivation by saying that we are trying to finish it as quickly as possible. We may enjoy walking or driving without any place in particular to go. In all of these instances, indeed, the termination of the activity may bring regret, rather than the satisfaction and release of tension usually considered to mark the end of a sequence of motivated behavior.

The fact that such cases tend to be neglected by motivational theorists does not mean that they present no problems. The questions to be discussed here concern the forces responsible for activity in the goal region.[2]

Lewin (4, p. 172 ff.) has opened discussion of this problem. What forces act on the person who is enjoying a happy state without much change—such as, for example, lying under a tree? Lewin suggests that the forces in question do not bring about a tendency to directed action, but only one which is opposed to leaving the region of positive valence.

The recognition of such a tendency, however, does not fully solve our problem. While it might account for *in*activity in a region of positive valence, it does not account for *activity* in such a region. In addition to resistance to leaving the goal region, forces must be postulated to account for behavior in this region.[3]

Phenomenally, the driving aspects of motivation are over when the individual has achieved his goal. How, then, can activity occur? We start with an examination of specific cases. In the following examples, it will be noted, the enjoyment may be referred either to the self or to the object. Correspondingly, the forces responsible for activity in the

[1] The writer wishes to express her thanks to her colleagues, Gertrude Baltimore, Florence R. Miale, and Dr. Irvin Rock, for stimulating discussions of this problem.

[2] Phenomenally, the satisfying experience of being together with what one has wanted is one of harmony between person and object. Another significant problem in this area is that of the nature of this harmony. What are the lacks of the person that the object meets? his aspects that the object "fits"? his properties that the object reflects or symbolizes? How much of each enters into the relation?

[3] Allport's theory of functional autonomy of motives (1) recognizes that activities may be engaged in for their own sake. It attempts, however, to account only for the selection of such activities, and does not discuss the forces responsible for them.

goal region may arise either from the needs of the ego or from the experienced properties of the situation. (Cf. **2**, **3**, **6**.) For the sake of convenience, we may describe the former as *ego forces*, and apply the term *field forces* to forces stemming from parts of the individual's psychological field other than his ego.

Ego forces are clearly primary in the case of driving for pleasure. The driver follows his desires (within the goal region, even though he is not striving to reach a goal), and the enjoyment is referred to himself. Again, a person's chief motivation in playing a game *may* be to win—an ego force—and the main enjoyment may derive from winning. While it might be maintained that we are dealing in such a case mostly with goal striving (to win), rather than with activity in a region of positive valence, even playing to win has usually in part the character of activity in a goal region; and the enjoyment of the winning, with its reference to the self, still needs to be accounted for.

In other instances it seems that field forces may play a considerable role. The enjoyment, as well as the forces responsible for the activity, are frequently referred to the object. In the case of reading the absorbing book, for example, it is not that I want to reach the end; rather the activity continues until the end stops it. The self, to borrow a phrase from MacLeod (**5**), accommodates itself to the object. Again, in the friendly conversation, a question elicits its answer, a feeling finds its response, an idea calls forth an elaboration. The forces responsible for its continuation *may* arise, not from extraneous motivation but, just as in the case of many other thinking processes, from the gaps and demands of the activity itself.

In both kinds of cases, it must now be added, whether enjoyment of activity in a goal region and the forces responsible for it are experienced as referring primarily to the self or primarily to the object, both ego forces and field forces are involved, and it is the relation between them that is important. The absorbing book, for example, must be absorbing to *me;* it must be within my comprehension and it must appeal to some interest or motivational system —conscious or unconscious—within me. In the case of playing the game to win, the game must be played according to its rules and within its specific limitations. The player's actions are called forth by the momentary situation as he perceives it. He plays a trump, for example, not because he is enamored of the particular suit, but because he sees the momentary distribution of cards as demanding this play.

In examining the above examples, one is struck by their similarity to instances of actual goal striving. That dynamic differences exist between the two kinds of motivated action is suggested, however, by the above-mentioned difference in outcome between them: satisfaction and release of tension in the case of goal striving as contrasted in many cases with regret when it becomes necessary to terminate activity in the goal region. The next step for investigation, then, seems to be to study the specific dynamic differences between the two kinds of cases.

It will be noted that the above difference in outcome is probably at present the most reliable criterion for distinguishing between these two kinds of activity. Enjoyment as such cannot be taken as a criterion, since it often pertains to striving itself as well as to activity in the goal region.

Another aspect of the problem of the forces at work in a region of positive valence is that of the relation between forces responsible for striving to reach a goal and those responsible for activity in the goal region. Here we see that

continuity is the rule. For example, both kinds of forces may operate simultaneously. In the case of the activity enjoyed with the friend, e.g., solving a problem together, the main forces may be those in the direction of completing the task (both field forces and ego forces or forces acting on the persons as a group); yet the work may go on in an atmosphere of closeness and humorous exchange that may give it also the character of activity in the goal region. Further, activities together with the friend—activities having at least partly the character of activity in the goal region—may result in changes in both persons. Not only will there follow changes in the details of the harmony between them, but new driving forces may also be initiated. For example, A may awaken new interests in B, or bring out dormant aspects of B, or show B he can do something B didn't know he could do, with corresponding changes in the motivated actions of B. Or else A may discover an aspect of B he didn't know about, e.g., an interest of B which A hopes to share, so that goal-directed forces act on A in relation to this quality. In all these cases the nature of the relation between A and B (and thus the nature of further activity in the goal region), as well as the subsequent goal-directed activity of A or B, will be altered.

Other changes in A or B may eliminate the harmony between them. For example, a person's taste in music changes; he is now a more complex A than he was when he enjoyed music B. B now seems trite to him and he moves on to something else. Or B may be badly played so that A cannot enjoy it. Such changes in the goal region set up forces driving the individual out of this region. A special case of this kind is that of satiation of activity in the goal region (cf. also 4, p. 141). In other cases the individual may leave the goal region, in the absence of satiation, under the influence of external pressures. The effects of insufficient early satisfactions on subsequent striving are well known.

It might be objected that it is misleading to view activity in the goal region as a special problem. Might we not describe such activity as a movement from one subgoal to another? Having made the friend, we are led to discover one new facet of the person after another. We move from one chapter of the book to the next. Each of these takes on the character of a subgoal, and the activity can be described in terms of the familiar forces driving the individual toward goals. Lewin (4, p. 172) has suggested that "consumption, such as eating . . . can . . . be viewed as a type of locomotion through subregions of consumptions, the character [valence] of which is progressively different. . . . It may be that every consumption has to some degree this character."

But he adds that there are cases in which the happy state is enjoyed without much change, which do not fit this description. The postulation of subgoals does violence to phenomenal experience in many cases of activity in a region of positive valence. It ignores the fundamental characteristic of these experiences, that of being activity in the goal region. Furthermore, the description in terms of subgoals takes into account only the *movement* from one phase of the activity to the next; it has nothing to say about the performance of any activity for its own sake.

The above remarks need to be qualified in one respect. Motives are so organized with respect to one another that the goal of one may be a subgoal with respect to a more inclusive need. Striving to succeed in a particular activity, for example, may be regarded as a chain

of motivated action in its own right, but may also serve more inclusive ego needs. In this sense, activity in the goal region may indeed be regarded in its aspect of movement to the next subgoal. But even in this case there probably remains the problem of activity in the subgoal region, as distinct from that of locomotion between subgoals. In addition, some of the cases cited above—for example aesthetic experiences and probably the case of the enjoyment of the friend—do not permit of this kind of analysis into subgoals on the path to more inclusive goals.

Summary

Attention has been called to a problem neglected by motivational theory, that of activity in the goal region. Examples have been discussed, and next steps for investigation have been indicated. The continuity between forces responsible for activity in the goal region and forces driving the individual toward goals has been suggested.

REFERENCES

1. ALLPORT, G. W. *Personality*. New York: Holt, 1937.
2. ASCH, S. E. *Social psychology*. New York: Prentice-Hall, 1952.
3. KÖHLER, W. *Gestalt psychology*. New York: Liveright, 1947.
4. LEWIN, K. *The conceptual representation and the measurement of psychological forces*. Durham: Duke Univer. Press, 1938.
5. MACLEOD, R. B. Perceptual constancy and the problem of motivation. *Canad. J. Psychol.*, 1949, **3**, 57–66.
6. WERTHEIMER, M. Some problems in the theory of ethics. *Soc. Res.*, 1935, **2**, 353–367.

(Received September 7, 1955)

Psychological Review
Vol. 63, No. 5, 1956

THE LOGICAL STATUS OF HULL'S PRINCIPLE OF SECONDARY REINFORCEMENT [1]

F. J. McGUIGAN

Hollins College

Hull's system is the most quantitatively advanced in psychological theory. But Hull himself recognized the need for the continuous development of the system. Referring to *A Mathematico-Deductive Theory of Rote Learning* (3) he says, "At one time a few of us worked out for a limited range of behavior a strict system to explore its possibilities. It is probably too early to do this on a large scale . . . [but] . . . such a task should before very long . . . [be attempted] . . . at least for the field covered by the present volume" (6, p. 3).

The accomplishment of this goal would be a major effort. The scope of this paper will be limited to a small part of the entire problem, representing a preliminary step toward developing a "strict system," i.e., a determination of the logical status of the principle of Secondary Reinforcement (Corollary ii), with the expectation that similar analyses could be offered for other principles in the system, e.g., the principle of Secondary Motivation (Corollary i).

In order to accomplish Hull's purpose, we might make use of what Carnap (1) calls the *postulational method.* This method has two phases: the *formalization phase* consists of the construction of formal systems, mainly the task of logicians and mathematicians; in the *interpretation phase* the empirical scientist chooses a formal system appropriate to his kind of data, and interprets the primitive terms of the formal system. These interpreted primitive terms are then combined, according to the formation rules of the formal system used, to form primary principles of the empirical system. From the primitive terms and primary principles, secondary terms and principles are derivable.

With reference to Hull's system let us ask: "Where might the empirical terms be found with which we can interpret the primitive (or possibly other) terms of a formal system?" Since Hull's principles concerned with reinforcement (particularly Postulate III, Primary Reinforcement, and Corollary ii, Secondary Reinforcement) are viewed by many as forming the foundation of his system, they constitute a good starting point for such an inquiry. Therefore, let us examine Postulate III and Corollary ii in an attempt to clearly delineate between primary and secondary terms, and between primary and secondary principles.

As a basis for further discussion, let us state the two principles: ***Postulate III:*** "Whenever an effector activity (R) is closely associated with a stimulus afferent impulse or trace (s) and the conjunction is closely associated with the rapid diminution in the motivational stimulus (S_D or s_G), there will result an increment (Δ) to a tendency for that stimulus to evoke that response." *Corollary ii:* "A neutral receptor impulse which occurs repeatedly and consistently in close conjunction with a reinforcing state of affairs, whether primary or secondary, will itself acquire the power of acting as a reinforcing agent" (6, pp. 5–6).

[1] The author wishes to express his appreciation to Clark J. Bailey for criticizing drafts of this paper, and to Mary Barbara Zeldin for checking the logic used.

Apparently, Hull considered that Corollary ii is logically derivable from Postulate III, possibly in conjunction with other principles (see footnote 4). He says, for instance, "These [corollaries] are placed in sequence with the postulates, each corollary following the postulate upon which it mainly depends . . ." (6, p. 4); it can be seen that Postulate III is followed by Corollary ii. On this same point he says, elsewhere, "The principle of secondary reinforcement . . . was used by us as a primary postulate in 1943. With the additional insight which has come through further research and study . . . it becomes evident that a derived or secondary principle may be deduced by a logical process from other primary principles which are available" (5, p. 26). He then informally discusses this reasoning by starting with Postulate III. The essentials of this reasoning are given in an example (5, p. 26). Briefly, Hull's notion is that the reduction of fear becomes attached to the stimuli and to the stimulus traces which are active at the time. The relaxation due to the reduction of fear then generalizes forward on those traces and gives rise to conditioned inhibitions, which can reduce proprioceptive stimulus intensity and so reduce the drive stimulus (a reinforcing situation). It follows that the stimuli which acquire the power of evoking the conditioned inhibition in this way become secondary reinforcing agents. However, an analysis of this reasoning suggests that *Hull has shown us how a stimulus can become a secondary reinforcing agent, and has not deduced Corollary ii,* formally or informally, either from Postulate III or from Postulate III in conjunction with other principles.

We shall now determine whether Corollary ii is a logical consequence of Postulate III. If this deduction *is* possible, then Corollary ii is in fact a secondary principle of the system, and the concept of secondary reinforcement would have the status of a secondary term. Otherwise, as things now stand, Corollary ii would have to be considered a primary principle.

In seeking this derivation, it is readily apparent that we are at a disadvantage, for although there is some similarity in the concepts expressed in the two principles, the words are noticeably different. Before we proceed, it would seem advisable to attempt to remove this disadvantage. In an effort to preserve the meanings which Hull intended, and to change the wording as little as possible, but also to make the logical relationships of the two principles more apparent, the following minimal modifications are offered:

Modified Postulate III

1. *If:* An effector activity is closely associated with a stimulus-afferent impulse or trace, and the conjunction is closely associated with a reinforcement; [2]

2. *Then:* there is an increment to a tendency for that impulse or trace to evoke that effector activity.

Modified Corollary ii

3. *If:* A receptor impulse is closely associated with a reinforcement; [3]

[2] "Reinforcement" maintains the same meaning that Hull intended in this context, that is, a state of affairs in which there is a rapid diminution in the motivational stimulus S_D or s_G (primary reinforcement and secondary reinforcement, respectively). However, it is not clear that Hull actually meant that there can be a rapid diminution in s_G. For instance, he says "Secondary reinforcement differs from primary reinforcement in that the former seems to be associated . . . with stimulation, whereas the latter seems to be associated with the cessation of stimulation, i.e., of the S_D" (4, p. 97).

[3] This modification of the initial conditions of Corollary ii is meant to increase the generality of the Corollary, for its concern is not

4. *Then:* there is an increment to a tendency for that impulse to acquire the power of acting as a reinforcement.

To determine whether this deduction can be made, we need to symbolize the modified Postulate III and modified Corollary ii (hereafter referred to as III and ii) in the notation of a logical system; for this purpose the calculus of propositions is used. We need then only insert an implication sign between the symbolized versions of III and ii, and compute the truth values of the ensuing conditional. Once the truth test has been carried out we can conclude that the deduction is either valid or invalid in this calculus (e.g., cf. 8, pp. 38–39). Before continuing with this procedure let us note that Proposition (1) presumably includes *all* stimulus-afferent impulses; therefore it is understood to include the receptor impulses referred to in Proposition 3; [3] hence we may write:

5. *If:* An effector activity is closely associated with a stimulus-afferent impulse or trace, and the conjunction is closely associated with a reinforcement [Proposition 1]; *Then:* a receptor impulse is closely associated with a reinforcement [Proposition 3].

To increase the possibility of the deduction we may now substitute Proposition 1 for Proposition 3 in Corollary ii, and inserting an implication sign between III and ii, according to the above specified procedure, we may write:

6. *If:* Proposition 1 implies Proposition 2, *Then:* Proposition 1 implies Proposition 4.

A truth-table test of Proposition 6, however, indicates that the deduction is invalid, i.e., ii is not derivable from III.[4] The reason for this appears to be that Proposition 4 is not contained in III. It would seem that there are two separate and distinct notions expressed in Propositions 2 and 4; first, through the conjunction of the afferent impulses or traces, effector activities, and rein-

restricted to *neutral* receptor impulses. It should be emphasized that these changes are minor. But even so, certain questions about the *statement* of the principles may occur to the reader. For example, would the *trace* of a receptor impulse, when it is associated with reinforcement, acquire the power of acting as a reinforcement? Hull would undoubtedly answer yes, but the proposition doesn't say so. What is the relationship between stimulus-afferent impulses and receptor impulses? We are assuming here that the class of receptor impulses is either identical with the class of stimulus-afferent impulses, or sufficiently more narrow to be included in the latter.

[4] A statement of Proposition 6 in the calculus of propositions is: $(1 \supset 2) \supset (1 \supset 4)$. The truth-table test is presented below, in which *T* means "truth" and *F* means "falsehood."

Possible Truth Values of the Atomic Propositions			Truth Values of the Molecular Propositions		
1	*2*	*4*	$(1 \supset 2)$	$\supset$	$(1 \supset 4)$
T	T	T	T	T	T
T	T	F	T	F	F
T	F	F	F	T	F
F	F	F	T	T	T
F	T	F	T	T	T
T	F	T	F	T	T
F	T	T	T	T	T
F	F	T	T	T	T

The implication sign between the two sets of parentheses indicates the major operation. It can be seen that where the truth values of the atomic propositions are "true," "true," and "false," respectively, the molecular proposition is "false." Thus, Proposition 6 is a synthetic statement, and not a tautology as would be required for the deduction in question to be valid. The author has tried out various other ways of stating III and ii, e.g., by further fractionating III, but the truth table tests always led to the same result.

It should be noted that Hull (5, p. 26) leaves open the possibility that certain other principles may be used for this deduction, though it is not apparent that any others are immediately relevant to the problem which we are dealing with. It is possible that Hull meant such things as the relevancy of Postulate II, which states the relationship between stimuli and stimulus-afferent impulses.

forcement, ***the stimulus-afferent impulse or trace tends to evoke that effector activity;*** and second, through the conjunction of the receptor impulse and reinforcement, ***the receptor impulse tends to acquire the power of acting as a reinforcement.***

It would appear that we have at least three courses of action. First, we might establish a further premise which relates Propositions 2 and 4. For example:

7. *If:* Proposition 2, ***Then*** Proposition 4.

That is to say, in effect, if a stimulus tends to evoke a response, then there is a tendency for that stimulus to act as a reinforcement. A truth-table test of Proposition 6 with the addition of Proposition 7 as a premise shows that ii is thereby deducible. Proposition 7 is certainly testable, and Hull would undoubtedly accept it (4, pp. 85–89), but its hierarchial status is undetermined. Until it is known that Proposition 7 is derivable from other principles of Hull's system, it should be carried as an additional postulate with "secondary reinforcement" as a primitive term, though this alternative suffers the disadvantage that the principle is fairly limited in scope.

Turning to the second, and the more parsimonious of the alternatives, we might restate III in a more general way, maintaining ii as a secondary principle. The most obvious solution is to generalize III so that Proposition 4 is included in it, e.g.:

If: An effector activity is closely associated with a stimulus-afferent impulse or trace, and the conjunction is closely associated with a reinforcement; ***Then:*** there will result an increment to a tendency for that impulse or trace to evoke that effector activity ***and for that impulse or trace to acquire the power of acting as a reinforcement.***

The third course would be to maintain both ii and III as separate primary principles, with "secondary reinforcement" as a separate primitive term. If we do so, we can take advantage of the relationship expressed in Proposition 5 as follows:

If Proposition 1, ***Then*** Proposition 2

but 5. *If* Proposition 1, ***Then*** Proposition 3

If Proposition 3, ***Then*** Proposition 4

therefore *If* Proposition 1, ***Then*** Proposition 4 [5]

We may, therefore, write:

8. *If* Proposition 1, ***Then*** Proposition 2; *If* Proposition 1, ***Then*** Proposition 4.

Which may be more succinctly written:

9. *If* Proposition 1, ***Then*** Propositions 2 and 4.[6]

Whichever of the three courses of action is pursued, a general conclusion can be reached which is expressed in Proposition 9; that is, that the stimulus-afferent impulse or trace referred to in III should acquire two powers: first of evoking the effector activity referred to in III, and second of acting as a reinforcement, though it should be remembered that in the third possibility it was necessary to use ii as a separate principle to derive these consequences.

An empirical test of Proposition 9 might take the form of an experiment conducted by Dinsmoor (2) in which he associated a stimulus (e.g., a light of short duration) with the bar-pressing response in a Skinner box. He then extinguished his rats under conditions that presumably allowed him to assess the functions of the stimulus as either a discriminative stimulus or as a reinforc-

[5] By the chain rule of inference (**9**, pp. 73–74).

[6] By 6e, "merging of implications" (**9**, p. 39).

ing stimulus. Dinsmoor found that the stimulus acquired both discriminative and reinforcing functions to about the same degree, and concluded that it is not necessary to distinguish between them other than in the operations that produce them. If we could regard the discriminative function of the stimulus as also an evoking function, then Dinsmoor's experiment would satisfy the conditions of Proposition 9, and the stimulus (afferent impulse or trace) could be tested as both an evoker of the bar-pressing response and as a reinforcer of that response. Skinner (**10**, p. 41) points out, however, that a discriminative stimulus is quite different from an eliciting (presumably evoking) stimulus, so that it would be necessary to modify Dinsmoor's design by eliminating the establishment of the stimulus as a discriminative stimulus and simply to associate the stimulus with the response (and with reinforcement) as specified by Hull's principles. But a serious methodological problem in conducting such an experiment should be mentioned. That is the problem of operationally distinguishing the evoking from the reinforcing powers of the stimulus-afferent impulse or trace. It may be that such a separation is impossible to make behaviorally with our existing techniques. The difficulty in making the very similar distinction between a discriminative stimulus and a reinforcing stimulus has been discussed by Dinsmoor (**2**) and by Keller and Schoenfeld (**7**, pp. 236–239).

Since Proposition 9 is unlimited in scope, it presumably means that *all* stimulus-afferent impulses or traces acquire the power of acting as a reinforcement and of evoking *all* the effector activities with which they are associated. One particularly interesting consequence of Proposition 9 would thus be that any given stimulus-afferent impulse or trace should, in addition to evoking the effector activity with which it was associated, reinforce that same effector activity (a sort of "reverberating circuit" affair). An empirical test of this consequence might also be made with an appropriate modification of Dinsmoor's design, though it is necessary here also to deal with the methodological problem cited above.

SUMMARY

In order to accomplish Hull's goal of developing a "strict" system for the field which he covers, it is necessary to distinguish, first, between his primary and derived principles, and second, between his primary and derived terms. As a preliminary step the relation between Hull's principles of primary and secondary reinforcement was investigated. It was determined that the latter is not deducible from the former. Three possible courses of action were presented: first, to add an additional premise, in which case his principle of secondary reinforcement is deducible; second, to generalize his principle of primary reinforcement so that his principle of secondary reinforcement is included in it; or third, to return to his position of 1943 and maintain both principles as primary ones (Postulates). Whichever course is adopted, one general conclusion can be reached: If the antecedent conditions of his principle of primary reinforcement are satisfied, a stimulus-afferent impulse or trace will acquire two distinct powers—of evoking an effector activity and of acting as a reinforcement. The empirical testing of this conclusion was discussed.

REFERENCES

1. CARNAP, R. *Logical foundations of probability.* Chicago: Univer. of Chicago Press, 1950.
2. DINSMOOR, J. A. A quantitative comparison of the discriminative and reinforcing functions of a stimulus. *J. exp. Psychol.*, 1950, **40**, 458–472.

3. Hull, C. L., Hovland, C. I., Ross, R. T., Hall, M., Perkins, D. T., & Fitch, F. B. *Mathematico-deductive theory of rote learning.* New Haven: Yale Univer. Press, 1940.
4. Hull, C. L. *Principles of behavior.* New York: D. Appleton-Century, 1943.
5. Hull, C. L. *Essentials of behavior.* New Haven: Yale Univer. Press, 1951.
6. Hull, C. L. *A behavior system.* New Haven: Yale Univer. Press, 1952.
7. Keller, F. S., & Schoenfeld, W. N. *Principles of psychology.* New York: Appleton-Century-Crofts, 1950.
8. Leblanc, H. *An introduction to deductive logic.* New York: Wiley, 1955.
9. Reichenbach, H. *Elements of symbolic logic.* New York: Macmillan, 1947.
10. Skinner, B. F. *The behavior of organisms.* New York: Appleton-Century-Crofts, 1938.

(Received October 26, 1955)

Psychological Review
Vol. 63, No. 5, 1956

AN EXAMINATION OF ONE ASPECT OF THE THESIS THAT PERCEIVING IS LEARNED

NICHOLAS PASTORE

Queens College

The question of the role of empirical determinants in perceptual processes has formed the historical proving ground for contending psychological theories (**4**). On the one hand we have Berkeley's theory with its exclusive emphasis on experience (**3**), and on the other we have *Gestalttheorie* with its minimization of experience (**13**). That not all basic issues have been resolved successfully is evident in the current trend that perceiving is the outcome of a learning process. We shall cite some examples from the literature which will serve the double purpose of illustrating current tendencies and the special meaning that we shall later attach to the empiristic thesis.

Miller and Dollard write that perception probably follows "the same laws of learning as do other responses, even though the responses producing these cues may possibly occur within the organizing centers of the brain" (**14**, p. 73, footnote). Hull has expressed himself along similar lines (**9**). Hastorf and Knutson write: "There is growing evidence that we learn to perceive, and that perception, like learning, is purposive" (**8**, p. 88). Peterson suggests that perceptions are *"learned reactions,* acquired by the usual methods of 'trial and error' learning" (**18**, p. 3). Snyder and Pronko write: "Our hunch is that some day psychologists will say that we learn 'to see' just as we learn to play hop-scotch, or poker, or read and write Arabic or English" (**23**, p. 111). Murphy writes that perceptual organization is "very largely the result of your lifetime experience in interpreting the impressions from your sense organs" (**15**, p. 144). In line with his "probabilistic functionalism," Brunswik writes of the "vicissitudes of probabilistic reconstitution of the object world" (**5**, p. 71).

Although cognition is sometimes used interchangeably or connected with perception (**2**, **16**), it is clear from the contexts of the cited statements that those aspects of perception which involve shape, size, and depth are intended. We shall exclude from consideration that aspect of perception which encompasses cognitive processes. Without attempting a definition of the term perception, it is clear that perceiving oscillatory movement in a continuously rotating trapezoidal figure is different in character from cognizing the meaning of a word. In the former, we might be dealing with a process which is stimulus bound, whereas in the latter, experiential factors certainly play a decisive role.

How shall we understand the contents of the quoted statements which reflect the empiristic tradition? The purpose of this paper is to define and evaluate an answer to this question.

Empiristic Meanings

Three of the several meanings which can be attached to the empiristic view of perception are as follows.

1. The structured cortical processes underlying perceptions, although autochthonously determined, require visual stimulation for their emergence or development. The requisite stimulation is here conceived in a maturational sense: it provides the opportunity for the growth of neural processes without

changing or altering their direction. For example, Riesen's experiments (**19**, **20**) suggest that visual stimulation is necessary for the proper development of retinal (and other) structures involved in perception; apparently the perceptions of his experimental animals, when they evolved, were not different from those of the control animals.

2. The second meaning is the same as the first except for the affirmation of the possibility of changing the structures themselves as a result of experience. Gibson expresses this meaning when he describes and rejects the following view:

> The formula being tested—that the perception of space is unlearned, meaningless, and acquires its meaning in the course of experience—makes an additional simplification which is far from perfect. It supposes that when a certain shape, let us say, gains a new meaning, the shape remains just what it was before. The texture, slant, color, contour, and other constituents of a thing are supposed to be unaltered by the gain in meaning. It may be apprehended differently but it is sensed the same as before, to use older terminology (**7**, p. 208).

Gibson then proceeds to cite various experiments which he interprets as supporting the view that "Color, size, form, sequence, and still other qualities of perception may unquestionably be affected by the past experience and attitudes of the observer" (**7**, p. 210).

Meanings 1 and 2 assume that the direction of perceptual processes is, at first, determined by intrinsic characteristics of the nervous system. Meaning 2 differs from meaning 1 in that it allows for a change in the cortical structures (and the corresponding perceptions) themselves. In order to illustrate this difference, we suppose that a distal elliptical stimulus is presented in the frontal plane to an individual; moreover, we suppose that the individual already has had whatever requisite stimulation is necessary to enable neural structures to develop. According to both meanings, such an individual perceives a specific form, perhaps that of an ellipse, a perception which is determined by autochthonous factors. Whereas in meaning 1 the individual can only perceive an ellipse, in meaning 2 the perception can be altered so that, with experience, the individual perceives a circle.

3. From contemporary discussions, it is quite apparent that still another meaning is intended by some of those who accept the empiristic approach. Prior to the advent of some learning process, a given stimulus situation evokes a variable perceptual process; as a result of learning, the given stimulus situation comes to evoke a specific perceptual process only. The view that the laws of behavior (those pertaining to the selection of an adaptive motoric response in trial-and-error learning) apply to "learning to perceive" lends support to the third meaning. Thus, according to this view, perceiving (which is regarded as a form of "response") is initially contingent or fortuitous in relation to a stimulus in the same way as a motoric sequence. Those who follow the "transactional approach" to perception postulate initial perceptual variability. For example, it is claimed that we must learn to see rectangularity even though the distal stimulus itself is rectangular,[1] since the typical retinal distribution of the stimulus is nonrectangular (**1**, p. 14). Acquaintanceship with the distal stimulus (through "action") leads to a perception which is in accordance with its actual characteristics. Presumably, *before learning*, the perceiver may see any one of an infinity of quadrilaterals which correspond to the infinity of external configurations to which a given retinal distribution is related (**11**, p. 3); *after learning*, the per-

[1] We are assuming that the distal stimulus is not perpendicular to the visual axis.

ceiver sees a rectangle only. Finally, the notion that spatial perception is learned also assumes initial variability of seen depth (or, sometimes, of its nonexistence). In stating one meaning that can be attached to this view Gibson writes:

> We know that the infant and young child ceaselessly explores his environment as his vision develops. Is it not likely that his visual impressions get their solidity and depth from their association with these movements? This is the argument for a motor theory of perception (7, p. 223).[2]

Presumably, either lack of motor experience or differently executed motor movements will lead to either a lack of or a variable perceived depth.

The present discussion will be confined to the third meaning that can be assigned to the proposition that perceiving is learned. Our thesis is that such a proposition must be rejected in view of the fact that an organism can learn various adaptive motor responses in relation to some stimulus situation. We shall assume, throughout, that an organism can learn only if a motor sequence is consistently reinforced in relation to some given aspect of the stimulus situation.

Initial Variability

What do we mean by an initially variable perceptual process? One answer is that the perceptual process issuing from the impact of some stimulus situation on the retina is random. Thus, on a phenomenal level, which we adopt for expository purposes, this would mean that a circular stimulus can elicit the perceptions of ellipses of varying degrees of eccentricities, rectangles, or any other geometrical form for that matter. Since no limits have been set to the central process, a circular stimulus may even elicit the perception of a cube. A similar example can be stated for depth perception. For instance, an object which is actually ten feet away from a subject can elicit a variety of depth perceptions, ranging from zero feet to some indefinitely large distance. Since this particular answer probably will be rejected on intuitive grounds, another and more realistic one will be stated, namely, that the perceptual processes elicited by a given stimulus are variable ***within a given range***.[3] Whatever form the perceptual process may have, it must be in accordance with the pattern of retinal stimulation. Thus, in a preceding example, a circular stimulus can only elicit perceptions of ellipses of varying eccentricities, and not perceptions of other geometrical forms.[4] Insofar as depth perception is concerned, however, the restrictive stipulation does not apply, since this would imply that objects in space intrinsically elicit appropriate depth perceptions.

In the foregoing we have essentially stated that the relationship between a given stimulus and the evoked perception is initially a fortuitous one; in order to enable our analysis to proceed, we assume a similar relationship between a given perceptual process and the elicited motor response.[5] We can clarify the latter assumption by considering a typical discrimination learning experiment from the point of view of the experimenter. A fixed stimulus situation—of which a certain aspect

[2] Gibson does not accept a positive answer.

[3] If our argument is valid when the restrictive stipulation is applied to meaning 3, it will certainly be valid when the restriction is removed.

[4] We note, in passing, that the effect of the restriction is to endow the nervous system with the capacity to respond in a characteristic manner to presented stimuli, a capacity which is not learned.

[5] A further justification for introducing a statement about the motoric system lies in the prevailing attitude that perceptions are learned through the organism's motor contacts with the environment.

may be denoted as positive, in that it leads to reward and the remaining aspect negative—is presented in a series of consecutive trials. The experimenter observes the change that occurs in the organism's response to the stimulus situation—a change that proceeds from diffuseness of response to the total stimulus situation to one of specificity to the positive aspect of the stimulus situation. Between the reception of the stimulus situation on the retina and the response, as noted by the experimenter, there are intervening processes which are not accessible to observation. One of these processes is perceptual, one which we have assumed to be variable. Moreover, for point of emphasis, we have also assumed that there is a variable relationship between an evoked perceptual process and the ensuing motor response. That is to say, a given perception may lead to any one of a number of motor responses. Of course, one could have assumed a unique relationship between perceptual processes and motor responses; the effect of this assumption, however, is to assert an intrinsic invariant relationship which an empirist might very well deny. In any case, whatever the attitude that can be taken toward this issue, the postulation of an intrinsic invariant relationship will not radically change our analysis of the thesis under consideration. However, we shall not examine the implications of such a postulation.

THE ARGUMENT AGAINST PERCEIVING AS LEARNED

We reject the proposition that perceiving is learned (in the sense of meaning 3) because, otherwise, the organism could never learn an adaptive response in relation to some visual situation. Two reasons can be offered for rejection. (*a*) In relation to a fixed visual situation, different perceptual processes (whether correct[6] or not) can elicit the consummatory response. (*b*) In relation to a fixed visual situation, the same perceptual process can elicit both consummatory and nonconsummatory motor responses. The implication of the foregoing reasons is that reinforcement of a motor response is not consistently related to the ultimately specific and correct perceptual process. The fact that the organism can and does learn consummatory responses invalidates that version of the empiristic view which was defined in accordance with meaning 3, since such learning is precluded.[7] We shall state our argument in an alternative way.

Let us define the set of perceptions which can ensue in relation to a fixed stimulus as $P(p_1^*, p_2, \ldots p_n)$ and the set of motor responses which ensue from each one of the perceptions as $M(m_1, m_2^*, \ldots m_n)$. The starred elements of P and M, respectively, denote the correct perception and consummatory response. Then, according to our exposition, any element of P can be linked to any element of M; of the *nk*

[6] By a "correct" perception we mean that perception which, after learning, prevails in relation to some stimulus. Returning to the example in which it was supposed that the perception of rectangularity (in relation to a distal rectangular stimulus) is learned, the perceiver may, initially, perceive any one of an infinity of quadrilaterals. After learning has been presumed to have occurred, the perceiver sees a specific kind of quadrilateral only, namely, a rectangle. For purposes of convenience, we label such a perception as ultimately comes to prevail as "correct," and all other perceptions as "incorrect."

[7] Our argument can be illustrated by considering an individual who must fit the various parts of a jigsaw puzzle. It would be impossible for the individual to execute the task if, in his perception, the contours of the patrices and the matrices underwent continuous and irregular transformations. Of course, the difficulty would be increased if the various parts also changed irregularly in seen depth.

possible linkages only one, $p_1{}^*m_2{}^*$, is correct and consummatory. In general, there are two types of linkages. In the first type, which illustrates *a* above, any element of P can be linked to $m_2{}^*$. In the second type, which illustrates *b* above, $p_1{}^*$ can be linked to any element of M. In effect, then, a stable $p_1{}^*m_2{}^*$ linkage cannot be formed.

We shall illustrate our discussion by two examples, one applying to form and one to depth perception.

Form perception. We have already mentioned that those who follow the "transactional approach" contend that the basic features of perception are learned. Among these basic features form perception is included. Thus, the contention that exposure to a "distorted room" from a certain point leads to a perception of a "normal-appearing room" by virtue of past experience can be subsumed under meaning 3. To simplify the discussion we shall consider a distal rectangular stimulus which is oriented in space so as to produce a trapeziform projection on the retina. Moreover, we consider an hypothetical individual who, in his first perception of the environment, is confronted with the rectangular stimulus. Such an individual, according to the *transactional approach* can have any one of an "infinity" of perceptions—the class of perceptions corresponding to the class of external configurations which can produce the given retinal distribution. Consider the way in which "action" can contribute to the formation of a perception of a rectangle. Suppose that our hypothetical individual is confronted with the tilted rectangular stimulus in a number of trials. Throughout this series of trials he will have random perceptions of the stimulus. In relation to each one of these perceptions the individual attempts to establish physical contact with the stimulus; those actions which lead to successful probing of the stimulus will then have a backward and reinforcing influence on the preceding perception. The "reinforced perception," of course, must be that of a rectangle, since this is the perception that ultimately will prevail. But, according to our analysis, a successful or an unsuccessful action can be matched with the perception of either a rectangle or a nonrectangular figure. Thus, on the basis of action alone no distinctive and uniform perception can emerge. It is possible that the individual can learn the appropriate probing movements (without, however, a distinctive emerging) since the same stimulus is presented in a series of trials. However, realistically, the individual may be confronted with other stimuli in the given series of trials; the stimuli are confined to those that produce the same retinal distribution. Thus, in line with the "reality" of the situation, in a first trial a particular rectangle is presented, on the next trial a larger rectangle is presented, and so on. It should be noted that as different rectangles are presented other changes must be made—changes involving distance of the rectangle from the subject and the angle of the rectangle with the visual axis. Consequently, in the reality situation, the individual could not even learn appropriate probing movements for any rectangle.[8]

Depth perception. Historically speaking, the notion that depth perception is learned was often stated in vague terms by its propounders (12, Ch. XX). The vagueness itself might have been partly responsible for the controversial status

[8] The writer realizes that the "transactional approach" has a wider scope than indicated here (for further comment, see 17). Moreover, it should be observed that the postulation of an infinity of possible perceptions throughout this discussion implies that the probability of occurrence of a particular perception is zero. In using the word "infinity" we follow the phrasing of the transactionalists.

of the notion. For example, the question of whether the phenomenal appearance of the spatially extended world was the outcome of a learning process, and the further question of whether different sequences of learning processes would lead to altered phenomenal appearances, do not seem to have been raised. Thus, in Berkeley's classical treatment of the problem (3) it is clear that although these questions were basic to his analysis, they were not explicitly phrased. In view of this ambiguity we feel entitled to define the notion under consideration in accordance with meaning 3. We shall illustrate this meaning and the argument of this paper in the next example.

The gibbon is a highly arboreal animal which can brachiate so rapidly through trees as to give the impression of running. Moreover, it is reported that this animal can span distances between tree limbs of up to 40 or 50 feet (6). Certainly there must be an occasion in the life of the gibbon when it must span a given distance for the first time; let us suppose that this given distance is 40 feet. In relation to this distance the gibbon must learn at least two things: the correct appreciation of distance and the correct application of a jump-force.[9] This example now parallels the situation described at the beginning of this section. Let us suppose that the gibbon is confronted with the same 40-foot distance in a series of trials. On the first trial, it "sees" the limb at a distance of 30 feet and then jumps; its jump carries it 50 feet. On the next trial, it "sees" the limb at a distance of 40 feet, but its jump carries it 30 feet. On the succeeding trial, the gibbon sees the limb at a distance of 20 feet and its jump successfully carries it to the limb, and so on. Obviously, in such a situation the gibbon could never learn to match a perceived 40-foot distance to a 40-foot jump. Or, more generally, it could never learn both the proper perception of distance and the proper application of force. Since we know that the gibbon successfully negotiates various distanced limbs, we must reject the underlying supposition that depth perception is learned. Our argument readily applies to other situations and to other animals. Moreover, we note that, for the gibbon, failure to span the 40-foot distance correctly the first time involves the risk of grave injury. Such a gibbon will not have much opportunity for learning through experience.

It can be contended that the distance of 40 feet is not actually new in the sense that the gibbon, having learned to perceive shorter distances, is able to transfer its depth response to the long distance.[10] To the writer this argument begs the question, since the capacity of the gibbon to so respond lies outside of learning.

The foregoing example suggests an evolutionary argument against the application of meaning 3 to perceptual learning; to ensure the biological survival of the organism it would be advantageous to have perceptual processes already pre-established in those situations of importance to the organism. Of course, on the human level, the situation may be different in kind from the one discussed here.

TWO ISSUES SPECIFIED

Two issues, implicit in the discussion up to this point, will be delineated.

1. It was assumed, in the examples that were presented, that the organism

[9] It is to be observed that in discussing the role of learning in depth perception, it is necessary to assume that form perception is already present.

[10] Our argument holds for *any* distance; however, with short distances, we probably cannot adduce the argument of injury to the animal.

is confronted by a fixed stimulus situation in a series of trials. Actually, in the organism's behavior in its environment, shifting, nonrepetitive stimulus situations prevail. In the last example, the physical relationships between gibbon and limb are variable as the gibbon locomotes; the distances, thicknesses, orientations in space, and angles of elevation of limbs do not remain constant. Moreover, background cues are most likely different from one limb to another. Consequently, stimulus variability increases the complexity of the organism's task in achieving perceptual learning.

2. For purposes of exposition, it was assumed that reinforcement and nonreinforcement facilitate or inhibit the recurrence of the corresponding perceptual processes. Thus, in the "distorted room" example, a single perception, that of a "normal-appearing" room, prevails in the given situation; other perceptions are either inhibited or replaced by the single perception of a "normal-appearing" room. The facts of learning, however, seem to be opposed to such an assumption.

In discrimination learning the organism can be trained to respond positively to one stimulus and negatively to another. Suppose that the two associated perceptual processes, corresponding to the positivity and negativity of the two stimuli, are either facilitated or inhibited. If the negative stimulus fails to elicit its corresponding perceptual process in the actual training situation then it should also fail to elicit the given process when it is presented as part of a stimulus compound. This implication is not in line with the evidence. Consider, for example, Woodbury's experiment in which a dog was conditioned to respond in a negative fashion to two stimuli when presented singly and positively to the same two stimuli when presented simultaneously (**24**). This result shows that the perceptual processes associated with the two negative stimuli could not have undergone inhibition as a consequence of nonreinforcement because the stimuli become positive when presented together. Afferent neural interaction, which is postulated by Hull to explain Woodbury's results (**10**, pp. 368 ff.), does not change the foregoing analysis, since before the afferent neural processes can interact they must be first elicited by their respective stimuli. An implication that can be drawn from this discussion is that facilitation or inhibition occurs on the motor response side, not on the perceptual side.

At least one phase of the present discussion can be checked empirically. It has been reported that a figure-ground relationship can be altered by a system of rewards and punishments (**22**). In their experiment, Schafer and Murphy presented a pair of profiles individually and tachistoscopically to subjects; one of the pair was rewarded and the other punished. When both profiles were combined, thus forming a single and more inclusive figure, it is reported that the rewarded profile was more frequently seen than the punished profile. Although the empirical result itself has been questioned (**16**, **21**), the interpretation proposed by Schafer and Murphy implies that the perceptual process associated with the negative profile undergoes inhibition. If this is actually the case, then it should be evident during the training period; that is to say, subjects should report increasing difficulty in seeing the negative profile.

SUMMARY

Of the three of several possible meanings that could be ascribed to an empiristic approach to perception, the one which involves an initially random perceptual process in relation to a given stimulus was selected for intensive

analysis. In order to enable the analysis to move forward, an initial fortuitous relationship was assumed between a perceptual process and the ensuing motor response. In general, the plan of the argument was to demonstrate that the assumptions of initial perceptual and motor random processes were incompatible with the fact of learned adaptive behaviors. The main effect of the argument, if valid, is to restrict further discussion of the thesis that "perceiving is learned" to the other meanings presented, or to some other meaning compatible with these two.

REFERENCES

1. AMES, A., JR. Visual perception and the rotating trapezoidal window. *Psychol. Monogr.*, 1951, **65**, No. 7 (Whole No. 324).
2. AMMONS, R. B. Experiential factors in visual form perception: I. Review and formulation of problems. *J. genet. Psychol.*, 1954, **84**, 3–25.
3. BERKELEY, G. An essay towards a new theory of vision. In *Berkeley: essay, principles, dialogues*. New York: Scribner's, 1929.
4. BORING, E. G. *Sensation and perception in the history of experimental psychology*. New York: D. Appleton-Century, 1942.
5. BRUNSWIK, E. *The conceptual framework of psychology*. Chicago: Univer. of Chicago Press, 1952. (*Int. Encycl. unified Sci.*, v. 1, no. 10.)
6. CARPENTER, C. R. A field study in Siam of the behavior and social relations of the gibbon (*Hylobates Lar*). *Comp. psychol. Monogr.*, 1940, **16**, No. 5.
7. GIBSON, J. J. *The perception of the visual world*. New York: Houghton Mifflin, 1950.
8. HASTORF, A. H., & KNUTSON, A. L. Motivation, perception, and attitude change. *Psychol. Rev.*, 1949, **56**, 88–94.
9. HULL, C. L. *A behavior system*. New Haven: Yale Univer. Press, 1952.
10. HULL, C. L. *Principles of behavior*. New York: D. Appleton-Century, 1943.
11. KILPATRICK, F. P. (Ed.) *Human behavior from the transactional point of view*. Hanover: Institute for Associated Research, 1952.
12. JAMES, W. *The principles of psychology*. Vol. II. New York: Holt, 1890.
13. KOFFKA, K. *Principles of gestalt psychology*. New York: Harcourt, Brace, 1935.
14. MILLER, N. E., & DOLLARD, J. *Social learning and imitation*. New Haven: Yale Univer. Press, 1941.
15. MURPHY, G. *An introduction to psychology*. New York: Harper, 1951.
16. PASTORE, N. Need as a determinant of perception. *J. Psychol.*, 1949, **28**, 457–475.
17. PASTORE, N. Review of *Human behavior from the transactional point of view*, F. P. Kilpatrick (Ed.). *Amer. J. Psychol.*, 1954, **67**, 379–383.
18. PETERSON, J. Aspects of learning. *Psychol. Rev.*, 1935, **42**, 1–27.
19. RIESEN, A. H. The development of visual perception in man and chimpanzee. *Science*, 1947, **106**, 107–108.
20. RIESEN, A. H. Arrested vision. *Sci. Amer.*, 1950, **183**, 16–19.
21. ROCK, I., & FLECK, F. A re-examination of the effect of monetary reward and punishment on figure-ground perception. *J. exp. Psychol.*, 1950, **40**, 766–776.
22. SCHAFER, R., & MURPHY, G. The role of autism in visual figure-ground relationship. *J. exp. Psychol.*, 1943, **32**, 335–343.
23. SNYDER, F. W., & PRONKO, N. H. *Vision with spatial inversion*. Wichita: Univer. of Wichita Press, 1952.
24. WOODBURY, C. B. The learning of stimulus patterns by dogs. *J. comp. Psychol.*, 1943, **35**, 29–40.

(Received October 31, 1955)

Psychological Review
Vol. 63, No. 5, 1956

PERCEPTION: IDENTIFICATION AND INSTRUMENTAL ACTIVITY[1,2]

FREDERICK H. KANFER
Washington University

Recent articles dealing with perception have introduced enough bewilderment to raise some question as to what behavioral phenomena are covered by the concept. If the various kinds of experimental procedures which are subsumed under the label are comparable, then the present divergence in defining perception could be avoided by a clear statement of criteria which define perceptual responding. On the other hand, if different operations and processes are described by the same term, then the danger of equivocation calls for a categorization of these studies in a clearer and theoretically more useful manner.

The perceptual process is a complex one. A broad definition would describe it as a behavioral sequence which is based on the contact of an organism with environmental objects and energies, and modified by the organism's history and its biological state at the moment of such contact. The major distinguishing characteristics of such a definition would be threefold: (*a*) A source of external stimulation is present; (*b*) the organism can make a discriminatory response to such stimulation, the nature of which is determined by its present biological make-up and earlier conditioning; (*c*) the final response in the sequence brings the organism into a different and often more adequate adjustment with respect to the recent change in the environment as it was discriminated.

It is the thesis of this paper that the wide variety of definitions and studies in perception is due not only to differential stress on various variables which operate at the same time, but also to differential interest in earlier or later responses in this sequence. Of two separate areas of research to which the term perception has been applied, it can be said that one has focused mainly on those responses that are correlated with external stimulation which *E* manipulates. The other has dealt with the consequences of these responses, i.e., with the behavior which is contingent upon the occurrence of some differential response to environmental stimulation.

The purpose of this paper is to suggest an S-R schema for description of perceptual processes which would allow the ordering of at least two classes of experiments on the basis of *E*'s focus and the extent of the contributions of each of the variables which are known to affect the behavioral sequence called perception. Such a schema may also help to reduce occasions in which we have differentiated as separate phenomena behaviors which can be better viewed as controlled by different quantitative contributions of various variables. Moreover, it is hoped that such analysis might be profitable in specifying research on the basic processes in-

[1] This paper was originated in conjunction with the author's investigation on verbal behavior which is supported in part by research grant M-963 from the National Institute of Mental Health of the National Institutes of Health, Public Health Service. Portions of this paper were read at the March 1956 meetings of the Southern Society for Philosophy and Psychology.

[2] The author wishes to express his appreciation to C. E. Sherrick for his helpful suggestions during the development of this paper and his critical reading of the manuscript.

volved in perception, and facilitate the integration of this area with other areas of psychology.

In dealing with perception, many investigators have been concerned with the behavior of the subject, which consists of a differential response to some complex properties of a pattern of sensory input, and which appears highly correlated with some dimension of the input. Such discriminatory responses have variously been discussed as sensory discrimination (17), identifying or naming responses (12), and recognition (1), or abstraction (22), and, in some experiments, as psychophysical responses. A second group of studies has dealt with the behavior of the subject during performance of an instrumental response under varying stimulus conditions which are usually under some measure of control by the experimenter. The inference is then made that the change in the pattern of sensory input acts as a discriminative stimulus, occasioning the instrumental response. In these studies there has been considerable discussion about the fact that frequently one fails to obtain a one-to-one correspondence between variations in the stimulus conditions and the observed responses.

Studies in this second class of experiments are exemplified by procedures in which *E* provides a connection between a discriminatory response and a subsequent motor response, often by means of verbal mediation. For example, *E* may instruct *S:* "When the second tone is louder than the first, press the key." Hence the instructions dictate the following sequence of events: (*a*) Two tones are given; (*b*) *S*'s discriminatory response to the tone occurs. In humans with verbal instructions, this differential response is likely to be a verbal one, i.e., "This tone is louder than the other." (*c*) Dependent on *b* above, *S* will then press the key, thus carrying out the motor response which is the dependent variable for *E*, or not press the key. The final observed behavior is a response which is contingent upon a certain differential response and triggered by it.

This contingency can best be demonstrated by first varying the strength of the verbal discriminatory response through pretraining in the absence of the instrumental response (the key-press). When the motor response is later introduced, it should differ as a function of the training. Proshansky and Murphy (19) demonstrated the effect of training such differential responses, in the absence of any overt measurable response, on the estimates of length of lines, and of weights.

Schematically similar, though vastly more complex, is the perceptual activity with which the clinical psychologist deals. In one projective technique, for example, *S* is instructed to tell a story about a picture. The clinician, from the story, then makes inferences about the differential responses which *S* may have made to various dimensions in the picture. He notes the events and characteristics of the stimulus picture to which *S* has reacted in the face of a large number of alternatives. He then assumes that *S* is most likely to react to many situations which may be similar to the present one in only one of its numerous dimensions, in much the same fashion. For example, "domination by the male" is judged to be the theme of a story given to a picture of a man and a woman. The clinician may infer that *S* perceives males as dominant in most situations which share only a few common elements with the given picture. The inference rests on the assumption that *E* has correctly chosen the variables which control *S*'s differential response to the picture from the story which he obtained.

As a device for ordering the experi-

mental variables which may control perceptual behavior, the two classes of research mentioned above can be schematized as follows (Fig. 1).

1. The first phase can be considered as a response to some change in the pattern of stimulation which impinges on the organism. This response is clearly not correlated with *all* dimensions of the sensory input. It is determined by (*a*) previous experience with similar physical dimensional patterns, (*b*) the limits of the organism's total capacities for responding at the given time and the limits of its receptors, and (*c*) the events antecedent to the present stimulation. Evidence of the contribution of each of these factors comes separately from experimentation in which the effect of the other variables has been minimized, although little is known of the relative importance of each in a complex situation. Studies on set and attention support the relevancy of antecedent conditions on the differential response. Studies on discrimination show that learning modifies the differential response, and comparative psychology offers evidence as to the limitations on discrimination, imposed by biological factors. If the effects of the variables under *a, b,* and *c* above are minimized, then the differential response should be contingent mainly upon the sensory input, and correlated with all or part of it. The mode of response, as indicated in Fig. 1, may be verbal, motor, autonomic, or a combination of these. In each case, however, it is a response specific to a complex stimulus pattern or part of it. Following Gibson and Gibson (12), this response will here be called an identifying response, and its occurrence will be observed as a function of any or all of the above variables. Under some conditions several conflicting classes of identifying responses may be available for the same stimulus complex. Under such conditions the specific variables under *a, b,*

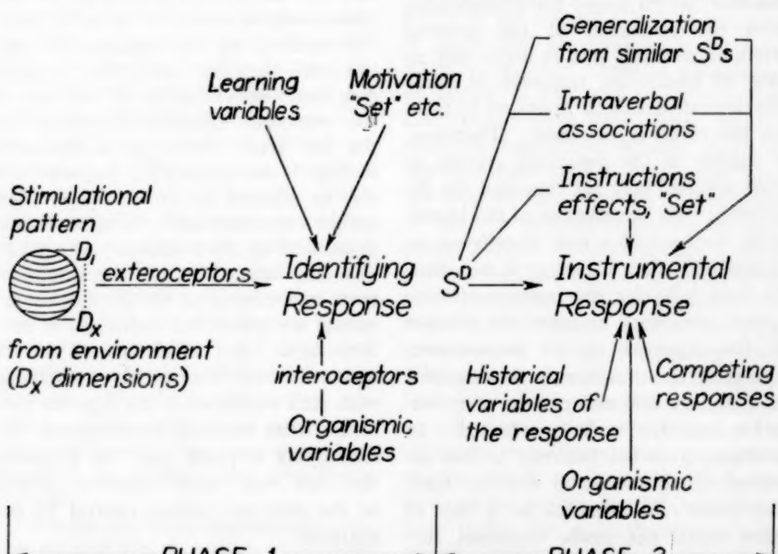


FIG. 1. The two phases of perception and some of the variables acting in each phase.

and *c*, above, such as drive state, previous reinforcement, generalization, etc., will determine the final response.

In some experiments the first phase is terminated by an overt identifying response, and its measurement constitutes the data of the study. In most studies, however, a second phase of the perceptual process can be discerned.

2. The identifying response serves as a stimulus for a series of other responses, motor or autonomic, which have been conditioned to it by past experience. These responses may be designated as instrumental responses. They have generally been described as perceptual responses which play a major role in bringing the organism to a new adjustment with respect to the recent change in the environment, as the organism has identified it in phase 1. The identifying response thus precedes and determines the activity which relates the organism to its environment. The subsequent instrumental responses, however, are no longer highly correlated with the properties of the external stimulus, and will tend to occur only as long as identifying responses or other discriminative stimuli produced by them set the occasion for them. Therefore, a change in the presented stimuli or their removal may not interrupt the *S*'s activity; only termination of the identifying response or a new identifying response will effect a change in behavior. In most behavior the instrumental response somewhat modifies the relation of the organism to its environment. Therefore the organism is now exposed to a slightly different pattern of stimulation, and the cycle is repeated. In addition, if verbal behavior is also involved, the instrumental response itself may serve as a stimulus for a host of other verbal and motor responses, further altering the conditions for the next identifying response.

DISCRIMINATION AND THE TWO PHASES OF PERCEPTION

The proposed schema for conceptualizing perceptual activity has been described in Fig. 1. We will now examine briefly the relationship between differential responding and the two phases of perception discussed above. Possibly the simplest type of situation which involves the differential response is one in which an organism is exposed to a stimulus object, and allowed to respond differentially to the pattern of energies received at any moment by its receptors, with whatever response is available. In the absence of controlled reinforcement, certain "characteristic" responses may then be observed whenever the object is presented. In addition to the stimulus object, however, several other factors also affect the probability of occurrence of any one response. If we apply here the notion suggested by Estes (9) and regard the stimulus object as a population of elements, only a sample of which is actually received by the organism on any one trial, then the identifying response may come under control of any one of the numerous stimulus dimensions on the first trial. Secondly, as indicated in Fig. 1, the identifying response will also be affected by generalization from earlier experience with the same dimensions, and by *S*'s momentary condition. The key operations for the establishment and control of the identifying response are systematic variations in sensory input, and reinforcement of responses which are closely contiguous with such variations in the stimulus pattern. After repeated presentations, the identifying response may be so modified that only highly selective aspects of the stimulus pattern control its occurrence.

On the infrahuman level the question of how a discrimination is established

has been given exhaustive attention. Various discrimination theories have dealt mainly with the occurrence of a previously specified (correct) response as contrasted with another class of (incorrect) responses and with situations in which: (*a*) discriminable differences in the stimuli are assumed to exist, and (*b*) training has produced specific overt responses in the presence of these differences. Generally, a response is chosen which markedly changes the animal's locus or his environment and is easily observable, e.g., jumping, bar-pressing or maze-turning. This *selected* response is then obtained with increased frequency if some contingency, such as food reward, is allowed to occur afterward. The response distribution is then functionally related to various experimental variables.

Discrimination learning, however, does not always proceed smoothly by sudden emergence of the correct response with progressively increasing frequency. One pitfall involves *E*'s designation of the dimension to which the animal actually responds. The controversy over color vision in the rat in the 1930s (**17**) nicely demonstrates the problem of isolating the stimulus variables affecting the identifying responses. When experimental conditions allow reinforcement of instrumental responses which follow different classes of identifying responses (i.e., responses under control of differing stimulus dimensions), the congruence between *E*'s stimulus manipulation and the distribution of instrumental responses is not accounted for unless we can establish the stimulus dimension to which the animal responds. This amounts to experimentally performing the operations here described as defining the identifying response.

From the continuity-discontinuity controversy and the discussion of presolution behavior has come the suggestion that other responses, apparently involved more directly with the "recognition" of the pertinent stimulus dimensions, also occur, especially before the discrimination is well established. Brogden (**4**) has summarized some of this evidence. VTE behavior, orienting responses, etc., have variously been observed. From these observations one can infer that some identifying responses to the external stimuli may occur early in learning, and may modify that response for which *E* has arranged elaborate measuring devices. Recently, Wykoff (**26**) has also pointed to another antecedent of the final response in discrimination studies, and showed the importance of the observing response which exposes the animal to the stimuli. Thus most experimental work with animals has concentrated its efforts on the final response in a discrimination process and its development, without more than casual mention of the responses which apparently occur early in training and deal with actual receptivity of the stimuli. The assumption is made that these drop out in later stages of training. Another alternative, of course, is that these responses are also conditioned and their briefer, smoother functioning in later stages of discrimination makes them less conspicuous to the observer. Since most of these identifying responses are generally not of interest to the experimenter, they are not set up to be obviously instrumental in manipulating the animal's environment. Therefore their measurement becomes extremely difficult, and consequently the usual operations in available discrimination studies do not allow a clear differentiation between the two phases of the perceptual process which we have described, except under highly special conditions. Admittedly, the generalization of the present schema to animal discrimination is somewhat more difficult to dem-

onstrate, though its applicability is suggested by the data which we have discussed.

With humans it is more feasible to isolate responses which are primarily controlled by external stimulation and which may "mediate" subsequent instrumental behavior. The chaining of responses during the perceptual process is most easily investigated if some verbal behavior is involved. We can often utilize the *S*'s ability for verbal report in order to verify the particular stimulus dimensions which have led to the differential response. In a simple discrimination we may present two objects to *S* and ask him to respond to a difference by stating it. Or we may facilitate the discrimination by indicating the dimension along which the objects differ, in our instructions, or by presenting a sample object and asking *S* "to find one like it" (**14**). In the psychophysical experiment, of course, such procedures are used; *S* responds to either of two stimuli by identifying the position of one object or event with respect to another along the same dimension, or he identifies the time period during which he can discriminate a single stimulus pattern which differs from the background stimulation. The latter situation is somewhat more restricted, since we may consider one of the two stimuli as the constant input (background stimulation) provided by events which are not controlled by *E*, and which he regards as a relative zero input for *S*.

The research methods described above represent only a small part of the field of perception. One might well distinguish these studies on the basis of the response which is scrutinized. The identifying response is usually the measured response. The question is often one about the discrimination potential of *S*. Even in such apparently clearly defined work, however, we occasionally encounter the possibility that the measured response is not determined by the specific stimulus which *E* presents but by other factors. For example, we find that, prior to the actual response, *S* may show a tendency to respond in a given way. This lack of independence of the identifying response of preceding response has been shown by Verplanck, Cotton, and Collier (**23**) in a study dealing with brightness discrimination at the threshold. Our knowledge of the effects of other learning variables on the psychophysical procedure awaits more detailed study.

In many human studies on perception the complexity of the analysis is further increased by the fact that *E* studies not the identifying response itself, but behavior subsequent to it. The area of concept formation provides a typical example of the complex perceptual process often discussed in relation to personality theory. In one approach, for example, *S* is given a series of blocks of different heights, colors, and shapes, and is asked to sort them into several piles. Here the overt response which is tabulated is usually the latency or the number of trials required to get a "correct" solution. There are several dimensions along which the blocks differ. As in our previous example, the initial response to the blocks is the identification of the pertinent dimensions. Several different identifying responses may be possible, and *S*s are often seen to shift from one to another before finally arranging the piles.

The identifying responses themselves, of course, may be partly determined by previous practice with similar tasks, and one or several of the dimensions may have "acquired distinctiveness" as cues (**15**). They may also be determined by physical limitations such as color blindness. Furthermore, they may be verbal, motor, or autonomic. The instructions, however, usually limit the

number of identifying responses which *S* will accept as "correct" by limiting the instrumental responses that are contingent upon them. This is done by specifying the number of piles to be sorted. The *Ss* are seen to differ in their behavior in several ways. Some may verbalize the pertinent dimensions along which the blocks differ and, based on this identification, proceed to sort the blocks. Other *Ss* will carry out the entire sequence of identifying and sorting the blocks on a verbal level first, then proceed on the motor level, while still others will proceed with the sorting task without any overt (verbal or motor) indication of the occurrence of the identifying response. In trial-and-error fashion, *S* may finally make an identifying response which leads to the motor response that fulfills the instructions. Such behavior is likely to lead to some reinforcement, which would then tend to strengthen the entire sequence preceding it, including the identifying response. The *E*'s attention, however, usually remains focused not on the identifying response but on the response which is a function of it, namely, the motor sorting. It is interesting to note that *E* can usually infer the identifying response after *S* has handled several of the blocks, although *S* has not yet completed the task. This inference is made about the identifying response from a response subsequent to it. The important point to be noted is that the overt response which later appears as datum is not directly a function of the external stimuli and their properties to which *S* has responded, but of the identifying response which has preceded the overt response.

That many experimenters are aware of the confounding of the two phases of perception is seen by their various efforts to provide for specification of the identifying response by instructions such as "Think aloud," or by postexperimental requests for introspection and verbalization about how *S* arrived at a solution. Although recent research has decreased its dependence on verbal reconstruction of the process, much work remains to be done in evaluating the variables which determine such different behaviors as described in the block-sorting example above, and in defining the limits of generality of such variables, both intraindividually and interindividually.

SOME IMPLICATIONS OF THE PRESENT SCHEMA

1. From our discussion it has been noted that there are two distinct sets of variables which may effect the final response in a complex perceptual situation (see Fig. 1). These are: (*a*) variables controlling the identifying response, and (*b*) variables controlling the association between the identifying response and the instrumental response, as well as the execution of the instrumental response itself. In view of the large number of variables which may act at several stages of the sequence, prediction from the stimulus input to the instrumental response has been poor, and ***post hoc*** explanations of the "meaningfulness" of the stimulus have been used to account for this failure. Such explanations have been aimed at providing knowledge about the intermediate step, the identifying response. The present schema clearly indicates how our success in prediction can be increased by knowledge of the identifying response. Assume that we have trained *S* to respond differentially and with great ease to area, brightness, and texture. Suppose further that he has three instrumental responses available for a class of objects of area A: R_{a1}, R_{a2}, and R_{a3}. Similarly, he has three other instrumental responses available for a class of objects of a given brightness B: R_{b1}, R_{b2}, and R_{b3}; and finally

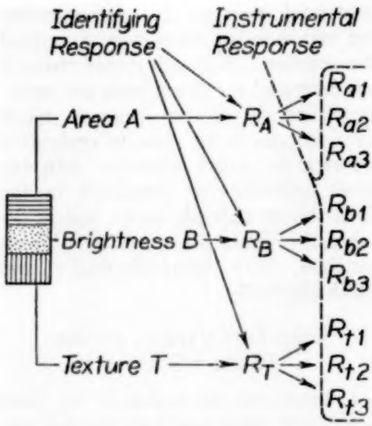


FIG. 2. An example of the availability of response classes in each phase.

R_{t1}, R_{t2}, and R_{t3} for a class of objects of texture T (Fig. 2). If we now present him with a stimulus object of area A, brightness B, and texture T, and if we further assume that no special conditions systematically favor the occurrence of any one instrumental response, then our best prediction for the occurrence of any of the available instrumental responses is that of $p = 1/9$. However, if the relative effectiveness of stimulus dimensions controlling the identifying response were systematically investigated, or if the dominant identifying response were established for an *S* from his behavior in other situations, it would then be possible to predict to which of the three dimensions *S* is most likely to respond.

For example, if such research suggests that A is the dimension to which a given *S* responds most frequently, we can now improve our prediction to read that R_{a1}, R_{a2}, and R_{a3} each will have the probability of occurrence of 1/3. Thus, the present conceptualization suggests one method of improving prediction in perceptual behavior. The example, of course, has been somewhat simplified, since other identifying responses may also be available to any combination of the three dimensions. Furthermore, the identifying response may be a function of the quantitative aspects of each dimension, and conditional discrimination may occur. Nevertheless, the increased complexity does not alter the approach suggested by the example.

In proposing the present formulation an obvious question arises about its accessibility to investigation. To date, support for ordering perceptual variables on the basis of the present schema has come mainly from two sources. Studies on stimulus predifferentiation (10) and transfer of discrimination (8, 14, 21) essentially utilize the operations here described for the identifying response in the pretraining phase, while subsequent training demonstrates the effect of differentially manipulating the identifying response on learning of later responses in the presence of the same external stimuli. Several mechanisms have been advanced to explain how these "pretrained" responses become effective (14, 21, 25). For the present schema unequivocal acceptance of any of these is not necessary. The experimental evidence resulting from all of these formulations, however, suggests the two-phase model of perception which would stress the linkage of responses to sensory stimulation with subsequent behavior, and which would order perceptual research into study of the components and the relationship between them. In many of the perceptual studies, the "pretraining" to experimental stimuli, such as pictures or meaningful words, encompasses *S*'s prelaboratory experiences with these and related stimuli. The effect of such pretraining is often given inadequate attention, and may result in misinterpretation of the data or faulty control of a significant variable.

The recent advent of information theory in psychology has stimulated some work on the variables controlling the identifying response, and promises to be of great value in perceptual research, since it provides one method for the treatment of multidimensional stimuli. Representative of this research are studies such as that of Bricker (3). In a study on the identification of redundant stimulus patterns, he attacked the problem of the effect of one dimension in the stimulus pattern on the identifying response. Eckstrand and Wickens (8) studied the effect of pretraining with color or form, or both, as relevant stimulus dimensions on the differential response, while Gregg (13) investigated the effect of varying the number of stimulus dimensions on reaction time in human discrimination.

2. From our formulation it is clear that the instrumental response in perception is more closely associated with a change in the physical relationship of *S* to his environment than is the identifying response. Therefore, reinforcement or nonreinforcement resulting from *S*'s changed relation with respect to the environment occurs temporally closer to the instrumental response than to the identification. Secondly, in human adults, identifying responses, even on the verbal level, usually are not made overtly. The act of perceiving is generally a "private" one. Thus, the identifying response is not as freely accessible to evaluation and modification by the social environment and therefore is not as universal in a given cultural setting.[3] In fact, in some cases, reinforcement for a perceptual behavior sequence may be obtained if the instrumental response is acceptable to the perceiver's fellow men, even though the identifying response may be quite unacceptable to them. Any clinician is familiar with such behavior in patients who obtain the "correct answer" or behave appropriately, often despite extreme deviation in their identification of a common external stimulus. From these considerations, our hypothesis may be stated as follows: In a given perceptual behavior sequence, the identifying response is changed or extinguished with greater difficulty than the subsequent instrumental response, since the latter may also result in reinforcement in situations in which there is no close correlation with the environmental stimulus pattern to which the former is made. While we know of no experimental work at present that may reflect on this hypothesis, it is easily testable. Our analysis also carries some implications for the study of verbal behavior. This area may best be approached by considering the multiplicity of variables which can act, either during initial formulation of *S*'s responses to environmental stimulation or when *S* attempts to explicate them to others in order to affect their behavior. Such a treatment of verbal behavior would bear similarity to Skinner's classification of tacts and mands (22).

3. It has already been stated that the instrumental response in our schema is contingent upon the identifying response and not directly on the stimulational pattern. Therefore, it may continue even after the stimulation pattern is changed, as long as the identifying response sets the occasion for it. Conversely, the identifying response may set off an instrumental response (even though the identification is not veridical) and maintain the instrumental response until a change in the discrimination occurs. Thus, our analysis provides for some explanation of the persistence of wrong hypotheses, as described by Blake and Vanderplas (2), and of the relative freedom of some

[3] In this connection, see Cameron's excellent treatment of socially shared and covert responses (5).

perceptual behavior from a strict dependence on minor changes in the stimulus pattern. It also suggests the possibility that maintenance of the identifying response in the face of some changes in the stimulational pattern may be involved in the persistence of instrumental responses which are no longer correlated with the stimulus which *E* presents. According to our formulation, except when other factors than learning are highly dominant, such behavior could be due to the more gradual extinction of the identifying response and the dependence of the instrumental response on it. Experimentally, one would expect extinction of perceptual responses to occur more quickly if one of the variables controlling the identifying response could be used to alter it directly.

4. On the assumption that organisms interact with their environment in numerous ways, it has already been suggested that the identifying response may occur on at least three different levels, the motor, the verbal, and the autonomic. If a stimulational pattern is responded to only on the autonomic level, for example, it would be expected that the instrumental response will be primarily dependent on it. Accordingly, the overt response may be made without a correlated report about its antecedents, in which case *S* is said to act without awareness of the stimulus for his reaction. While the stimulus information may be adequate for "autonomic identification," it may not be sufficient for a verbal response. On the basis of the instructions, *S* may then report: "I do not see it" or, "I see a blank screen," despite the fact that he gives a GSR. Some studies on subception (**16**) in which previously punished word stimuli are presented below threshold for verbal recognition seem more parsimoniously treated in this way than by invoking preperception or other censoring devices. It also is possible that two identifying responses are made to different aspects of the same stimulational pattern, each in a different mode of expression. To borrow an analogy from information theory, one might speak of contradictory information, transmitted over two different channels (**18**). The resultant instrumental behavior should show some lack of integration, since components of it are partly determined by two different identifying responses. One example might be the identifying response of the listener when speech and gestures of the speaker "do not match." The clinician who observes tears in the eyes of a patient who tells him that he is feeling wonderful is in such a situation. Several questions of interest are raised by these speculations. What factors determine the pattern of stimulus dimensions which is selectively responded to? For example, does training have an effect on the dominance of the channel to which the response is made? To what extent do motivational variables affect the identifying response? Are these variables important in determining the mode of responding?

5. The use of verbal behavior in identification allows for considerably sharper discriminations than the vague, diffuse stimulational patterns provided by our interoceptors. Furthermore, through language the explication of identifying responses in communicable form allows for establishment of instrumental responses which fall under social control. Dollard and Miller's stress on labeling in psychotherapy stems from a similar rationale (**7**). Only after the identifying response of the patient toward a given situation has been established in verbal form can he be taught to make adequate instrumental responses which are more generally adaptive. The therapist, by teaching the patient to respond more effectively to pertinent cues from his environment,

provides naming responses which can then be related to instrumental activity. In a sense, therapists teach their patients to "see things differently," i.e., to respond to different stimulational patterns than before and to do so in a verbal mode of response. The role of such naming responses has recently been studied in various experiments (**6**, **20**, **21**), and promises to shed light on this major component of human perceptual behavior, i.e., language behavior.

6. A number of recent studies have dealt with perception of ambiguous stimuli. Let us define ambiguity with respect to the number of responses available for a given set of stimuli. Given a specific stimulus pattern to which *S* has available one class of identifying responses which are at high strength under these conditions, the effect of other variables which control the identifying response will be minimized, and a veridical perception is reported. As the stimulus input becomes less highly correlated with an identifying response of high strength, alternate responses may occur, leading to instrumental responses which may or may not bring the organism into more advantageous relationship to his environment. Such a process might be described as "checking reality" or "testing hypotheses." If the stimulus information is reduced, more patterns in the stimulus input may be discriminated and a greater number of responses to them becomes available. Such a reduction in stimulus information can be obtained by decreasing exposure time, omission of cues from familiar objects, reduction of illumination, or offering cues which normally do not occur together in familiar settings. Under these conditions, the competition of responses should be greater and the effect of variables which we have roughly classed as organismic, learning, and motivational (see Fig. 1) may have greater weight in determining the response.

On the basis of our analysis there is, however, a second source of "misperceptions." While we may call the former misrecognition, there is also the opportunity for ambiguity in the second phase. Despite veridical recognition, there may occur an instrumental response which is not highly correlated with any one identifying response. In studies with taboo words, questions about this point of variability appear in connection with withholding verbal report of the stimulus. Thus, there are two sources of variability in perceptual processes when we present stimulational patterns for which there is no widely accepted or highly trained response. First, we may fail to obtain agreement in the identification of the pattern when a low correlation exists between the stimulus input and an identifying response of high strength. Secondly, the behavior subsequent to recognition may be controlled by variables which are not explicit or not highly uniform in a population. The projective tests have capitalized on both sources of variability, though the distinction is not often made.

The traditional Rorschach procedure provides an interesting reflection on our analysis. Here, after *S* gives some verbal report about the blot, *E* follows up with an inquiry. In this inquiry he asks for an explication of the stimulus dimensions to which *S* has responded, i.e., whether form, color, or shading was used. The assumption is made that *S* will show selectivity with respect to the dimensions in the stimulus to which he tends to respond, and that such selectivity is related to personality variables. The Rorschach inquiry, then, is one attempt to make overt the identifying response, as we have here described it. Other levels of Rorschach interpretations, such as con-

tent analysis, usually reflect on the instrumental response which follows recognition, and are separately related to personality variables.

One other illustration of the application of our analysis can be cited in relation to an interesting series of studies by Weisskopf (24). She proposed the use of a Transcendence Index as a measure of the number of comments which go beyond mere description of pictures on the TAT. In this study, contrary to standard TAT practice, *S*s were asked to describe *what they saw*. Her failure to obtain uniformity in description attests to our earlier suggestion that even the first phase of the perceptual process, the recognition and identification of a stimulus pattern, is not completely determined by the sensory input alone. In the usual TAT procedures, of course, the process also includes phase 2, since *S* is asked to tell a story based on the card; i.e., the final response is one which is subsequent to the identifying response, and additional opportunity for the operation of personality variables is thereby given. The role of the identifying response may profitably be studied here by differential reinforcement of card description and study of the changes in subsequent story-productions.

7. The concept of set has been an especially important one in recent theorizing about perception. Its application has been so widespread, however, that numerous referents can be found for this term. Generally, it has been implied that set may have its effect, be it facilitating or inhibiting, either with respect to selective receptivity (sensorial) or with respect to the following motor response. In our schema those variables which have been subsumed under *set* can be separated into: (*a*) those affecting the identifying response and (*b*) those affecting the instrumental response. Experimental analysis of the effect of set or attitudes could then be further facilitated by investigations which focus at each point in the perceptual sequence.

SUMMARY

The present paper suggests a conceptualization of the perceptual process which distinguishes two distinct phases in perception: (*a*) an initial response related to some dimensional pattern in the environmental stimulation; and (*b*) an instrumental response which is contingent upon the former. This formulation is presented primarily with the intent of ordering available data on perceptual processes, and suggesting further areas of research which would systematically evaluate the effect of various variables on the components of the complex perceptual process. In discussing the implications of the formulation, it was indicated what meaningful questions could be raised on the basis of the view which is presented here. Separate investigations of these steps should further our knowledge about two major questions posed in this area, namely, (*a*) how *S* comes to respond to certain selected aspects of his environment; and (*b*) what determines his behavior subsequent to and partially dependent upon his appraisal of the environment.

REFERENCES

1. BINDER, A. A statistical model for the process of visual recognition. *Psychol. Rev.*, 1955, **62**, 119–129.
2. BLAKE, R. R., & VANDERPLAS, J. M. The effect of prerecognition hypotheses on veridical recognition thresholds in auditory perception. *J. Pers.*, 1950, **19**, 95–115.
3. BRICKER, P. D. The identification of redundant stimulus patterns. *J. exp. Psychol.*, 1955, **49**, 73–81.
4. BROGDEN, W. J. Animal studies of learning. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951.
5. CAMERON, N. A. *The psychology of behavior disorders*. Boston: Houghton Mifflin, 1947.

6. CANTOR, G. N. Effects of three types of pretraining on discrimination learning in preschool children. *J. exp. Psychol.*, 1955, **49**, 339–342.
7. DOLLARD, J., & MILLER, N. E. *Personality and psychotherapy.* New York: McGraw-Hill, 1950.
8. ECKSTRAND, G. A., & WICKENS, D. D. Transfer of perceptual set. *J. exp. Psychol.*, 1954, **47**, 274–278.
9. ESTES, W. K. Toward a statistical theory of learning. *Psychol. Rev.*, 1950, **57**, 94–107.
10. GAGNÉ, R. M., & BAKER, KATHERINE E. Stimulus pre-differentiation as a factor in transfer of training. *J. exp. Psychol.*, 1950, **40**, 439–451.
11. GIBSON, J. J. A critical review of the concept of set in contemporary experimental psychology. *Psychol. Bull.*, 1941, **38**, 781–817.
12. GIBSON, J. J., & GIBSON, ELEANOR J. Perceptual learning: differentiation or enrichment? *Psychol. Rev.*, 1955, **62**, 32–41.
13. GREGG, L. W. The effect of stimulus complexity on discriminative responses. *J. exp. Psychol.*, 1954, **48**, 289–297.
14. KURTZ, K. H. Discrimination of complex stimuli: The relationship of training and test stimuli in transfer of discrimination. *J. exp. Psychol.*, 1955, **50**, 283–292.
15. LAWRENCE, D. H. Acquired distinctiveness of cues: II. Selective association in a constant stimulus situation. *J. exp. Psychol.*, 1950, **40**, 175–188.
16. LAZARUS, R. S., & MCCLEARY, R. A. Autonomic discrimination without awareness: a study in subception. *Psychol. Rev.*, 1951, **58**, 113–122.
17. MUNN, N. L. *Handbook of psychological research on the rat.* Boston: Houghton Mifflin, 1950.
18. OSGOOD, C. E., & SEBEOK, T. A. (Eds.) Psycholinguistics: a survey of theory and research problems. *J. abnorm. soc. Psychol.*, 1954, **49** (No. 4, Part 2).
19. PROSHANSKY, H., & MURPHY, G. The effects of reward and punishment on perception. *J. Psychol.*, 1942, **13**, 295–305.
20. ROBINSON, J. S. The effect of learning verbal labels for stimuli on their later discrimination. *J. exp. Psychol.*, 1955, **49**, 112–114.
21. ROSSMAN, I. L., & GOSS, A. E. The acquired distinctiveness of cues: the role of discriminative verbal responses in facilitating the acquisition of discriminative motor responses. *J. exp. Psychol.*, 1951, **42**, 173–182.
22. SKINNER, B. F. *Science and human behavior.* New York: Macmillan, 1953.
23. VERPLANCK, W. S., COLLIER, G. H., & COTTON, J. W. Nonindependence of successive responses in measurements of the visual threshold. *J. exp. Psychol.*, 1952, **44**, 273–282.
24. WEISSKOPF, EDITH A. A transcendence index as a proposed measure in the TAT. *J. Psychol.*, 1950, **29**, 379–390.
25. WICKENS, D. D., & BRIGGS, G. E. Mediated stimulus generalization as a factor in sensory pre-conditioning. *J. exp. Psychol.*, 1951, **42**, 197–200.
26. WYCKOFF, L. B., JR. The role of observing responses in discrimination learning. *Psychol. Rev.*, 1952, **59**, 431–442.

(Received August 10, 1955)

Psychological Review
Vol. 63, No. 5, 1956

PERCEPTION AS SUBSTITUTE TRIAL AND ERROR

DONALD T. CAMPBELL

Northwestern University

The Darwinian concept of random variation and natural selection through differential survival provides a general model for the deterministic description of adaptive phenomena. Recently this model has seen increased application to the adaptive phenomena described by psychologists as problem solving or learning. It is the purpose of this paper to suggest the application of the model to a third level, that is to say, to the adaptive behavior demonstrated in carrying out or executing a well-learned habit. Such an application seems to necessitate the conceptualization of sensory and perceptual processes implied in the title.

The early comparative biologists and evolutionists assembled impressive evidence of the adaptive fit between organismic structure and environmental possibilities. To explain such fit, three principal alternatives were available. The first involved the detailed a priori planning of a prescient deity. The second involved appropriate or corrective structural modifications based on experience with the environment in question. But this model also involves prescience, however modest and distributed, in that the organism somehow 'foresees which modifications will fit better. Where the Lamarckian notion of inheritable habituation could be applied, some plausibility might be gained. But most instances of adaptive fit could not thus be explained. The third model was the Darwinian theory of natural selection. For this model, unlike the second, the modifications or variations are blind, are random, are individually nonappropriate, are not of the order of corrections. But by chance there do occur those which provide better fit, and these survive and are duplicated. While Darwinian theory of evolution has undergone considerable elaboration and modification, and while there has been disagreement as to the mechanism and the magnitude of the variations involved, his basic model of natural selection is uniformly accepted today, and stands as one of the great conceptual achievements of the 19th century. In its abstract or formal aspects, it is a model which may be applied to other adaptive processes, or other apparently teleological series of events in which modifications seem guided by outcome.[1]

[1] Recently this model has been applied to embryological morphogenesis and differentiation. While this application is yet incomplete, it is worthy of psychology's attention, inasmuch as the "organizing fields" of the experimental embryologists have provided an influential model for field theory in psychology. The problem has been to explain how genetic control is exercised to account for growth and differentiation into adult form. A tempting early model involved the notion that the genes of the germ plasm, containing the total blueprint for the adult organism, became distributed in the process of mitosis, so that each somatic cell contained only those genes appropriate to its body part. This model required that the genes collectively contain an almost incredible detail, with every phase of growth predestined at a cell-by-cell level. The model has been invalidated by the experimental facts of wound healing, regeneration, and, most dramatically, by the transplantation experiments done on amphibian embryos. These show that, up to a certain stage, transplanted cells will develop in conformity with the locality in which they are placed, rather than in terms of the locality of origin. From such experiments have come concepts of organizing fields with excitatory and inhibitory gradients. Such concepts provide useful summaries of the facts, and may be stated in mathematical form, but are un-

SELECTIVE SURVIVAL IN PROBLEM SOLVING OR LEARNING

A second level of organismic fit to environment is to be found in the behavioral modifications that occur within the life span of a single organism. Thus, a rat's behavior comes more and more to "fit" the maze, to reflect the environmental possibilities. The consistency of the trend toward fit and the general goodness of fit achieved have tempted some writers to teleological explanations, or to explanations involving prescience on the part of the animal in small or large degree. Ashby (3) and Pringle (22) have explicitly proposed that the blind variation and selective survival model be applied to adaptation at the level of animal problem solving and learning.[2] There are differences in their applications. Ashby attempts to start from scratch, and to imitate initially a protozoan as simple as Jennings' *Stentor*. Pringle starts at the level of mammalian brain function, and suggests a model in which the selective survival of neural inputs among central nervous system resonance patterns accounts for learning. The present paper is most influenced by Ashby. To quote from his introduction to *Design for a Brain:*

> The work also in a sense develops a theory of the "natural selection" of behaviour-patterns. Just as in the species the truism that the dead cannot breed implies that there is a fundamental tendency for the successful to replace the unsuccessful, so in the nervous system does the truism that the unstable tends to destroy itself imply that there is a fundamental tendency for the stable to replace the unstable. Just as the gene pattern in its encounters with the environment tends toward ever better adaptation of the inherited form and function, so does a system of step- and part-functions tend toward ever better adaptation of learned behaviour (3, p. vi).

Ashby has designed and built a "Homeostat" which achieves adaptation at the levels of "stability" and "ultrastability," as meticulously defined. Essential to this machine are uniselectors or stepping-switches designed to provide a series of random changes in certain internal circuits. This device is activated whenever a swinging magnet is deflected out of its normal center range. Such deflection may result from interferences with any part of the machine's circuitry, such interferences represent-

satisfactory in not specifying how the control comes about. Spiegelman (26) has suggested that the natural selection model be applied here. In simplified form, he sees each somatic cell as containing the full complement of genes. The genes are conceived of as potential initiators of self-duplicating enzyme-formation processes, each gene producing a characteristic enzyme, which in turn determines a characteristic cellular structure. Which genes or enzymes dominate depends upon the character of the immediate surroundings, both within the cell and externally. The surroundings provide the setting for the selective propagation of some self-duplicating enzyme complexes at the expense of others with which they compete for energy and space. Amputation, transplantation, etc. change these surroundings, and thus the patterns of dominance. The pattern of growth is modified by changing the selecting conditions. While Spiegelman has not yet extended his model to describe fully the mechanism of embryological development, the direction of theoretical development provides a promising illustration of the application of the natural selection model to a dramatic instance of a seemingly teleological phenomenon.

[2] The parallel between natural selection and trial-and-error learning seems obvious, once pointed out, and yet is an important and profound intellectual achievement. It is interesting to note that Lloyd Morgan and E. L. Thorndike, who were imbued with the Darwinian heritage and were proponents of the trial-and-error doctrine, seemed to miss it, or at least did not make it explicit. Ashby and Pringle, both participating in the intellectual ferment out of which cybernetics and information theory have developed, seem to have noted the parallel independently. It is not here intended to establish precedence for the notion, however. Young (38) also presents it, although less clearly. And Lewin (16, p. 66) and Asch (2, pp. 96–98) both raise the parallel, if only to reject its applicability, apparently independently of other sources.

ing the impingement of the environment. The stepping-switches continue to emit the random changes as long as the swinging magnet is out of center, being stopped only when centering is again achieved, leaving the internal wiring as it is at that moment. In this sense the machine learns, or solves problems, introducing adaptive innovations into its own structure.

In terms of the rudiments of the general selective survival model, habit formation would be based upon (*a*) random variation of emitted behavior, (*b*) *selective* survival of certain variations, and (*c*) retention and duplication of surviving variations. In terms of conceptual traditions in psychology, this translates into a random trial-and-error learning model. Accepting the general correctness of the model may be made more palatable by noting that most contemporary learning theories contain a random trial-and-error component. While this is most obvious in theories of the Thorndike tradition, as in Hull, Skinner, and especially Guthrie, it is also true of the more Gestaltish theories of Tolman and Meier, as has been pointed out elsewhere (**6**). The major unsolved problems lie in the mechanisms of selection and retention, and these problems are formidable. But for the purposes of the present paper it is the randomness that needs emphasis.

The term ***random*** is an unsatisfactory one. Contained in the folk terms "random," "chance," "haphazard," "fortuitous," and the like are a number of connotations which have been seized upon and formalized in mathematical and scientific thinking. These are essential in the selective-survival model to varying degrees, and the absence of strict randomness on some points becomes a source of confusion. Speaking of the device for changing internal circuits, Ashby says:

> The values of the components . . . were deliberately randomized by taking the actual numerical values from Fischer and Yates' Table of Random Numbers. Once built on to the uniselectors, the values of these parameters are determined at any moment by the positions of the uniselectors. Twenty-five positions on each of four uniselectors . . . provide 390,625 combinations of parameter values (**3**, p. 96).

The settings or "trials" of the machine when disturbed approach randomness in all senses of the word. First, all possible settings are equiprobable. However, the device would still work even if this feature were considerably modified. Indeed, to refer to the evolutionary paradigm, it is known that the mutation frequencies for different genes vary widely. That responses differ greatly in likelihood of appearance is not an essential deviation from the model, as long as responses vary and continue to change in the face of distress.

Second, the settings or responses are independent of each other—the likelihood of a given response's occurring is not affected by the prior responses. This requirement can be considerably breached and Ashby's machine would still work, and indeed it seems unlikely that it holds strictly for the Homeostat as constructed. Furthermore, granted the frustration tolerance which the machine seems to have, it could still adapt if the uniselectors ran through the 390,625 possibilities in an orderly, systematic fashion. Mechanically it is probably simpler and more generally effective to construct the machine as Ashby has done, but it is not essential. Since there are among the 390,625 possibilities many with equivalent effects, or since for any given stress there are numerous potential resolutions, and since in a systematic order the adjacent settings would tend to have similar effects, the systematic approach would probably result more frequently in long

runs of trials without a solution, even though the modal solution time might not be greater.[3, 4]

Third, random connotes that the settings or trials are uncorrelated with the stress, or the stimuli that set them off. While this can be abrogated considerably without incapacitating the mechanism, a high correlation between antecedent conditions and response obviously eliminates problem solving, discovery, or innovation. Fourth, random connotes that the occurrence of the trials individually is uncorrelated with the solution, and in particular that specific correct trials are no more likely to occur at any given point than specific incorrect ones. This requirement seems essential. Insofar as it is breached in empirical observations of trial-and-error learning, it is under conditions of a joint correlation of response and solution with the instigating conditions, and represents prior learning or other prior sources of partial information. Genetically determined response hierarchies to stimuli could in this sense reflect an evolutionary accumulation of partial knowledge about probably appropriate responses. "Blind variation" or "blind trial and error" are better phrases perhaps, but "blind" used in this sense has metaphorical implications which overlap with the main theme of this paper. The term "nonprescient" is perhaps most appropriate.

A fifth and essential connotation of random or "blind" is the eschewing of any notion that a subsequent response is a "correction" of the preceding one, or makes use of the direction of error of the earlier ones. Such notions introduce imputations of prescience which a deterministic model must avoid, unless a specific mechanism for such prescience be introduced. The settings, trials, or responses succeed each other in a blind or random fashion, the subsequent responses being no more appropriate than the prior ones, except by chance.

[3] In the evolutionary setting of gene mutations, this aspect of randomness appears essential. If mutations tended to occur in a specific order, with the same mutation first, and if the great bulk of mutations are ineffective or lethal in results, then the possibility of species extinction without adaptive innovation becomes great.

[4] For the trial-and-error behavior of living organisms, there is another relevant consideration. A high degree of predictability in locomotor search behavior is a source of weakness is predator-prey relationships, as is emphasized in the theory of games.

ADAPTIVE FIT IN THE EXECUTION OF HABITS

A third level of adaptive fit of organism to environment occurs in the execution of habits. Not only does the rat manifest adaptive fit in solving or learning the maze, but in addition it usually manifests adaptive fit every time it runs the maze. Usually it has *not* learned a specific series of muscle contractions which run off regardless of the environment. Rather, it runs in conformity with the location of walls and passageways. Start it off on a different heading, or on a different foot, and its behavior still fits. Shift the usable mode of locomotion from wading to swimming, and although double the number of leg movements is required, the behavior still fits (17). While not all learned responses manifest this character, it seems that most of them do (6). This adaptive fit in the execution of habits has been called *molar* behavior by Tolman (30) and others. It represents *acts,* not *movements,* or *advertent* rather than *inadvertent* responses in Guthrie's terminologies (10). Brunswik (5) has designated it as consisting of *distal* responses rather than *proximal* ones. It is clearly the kind of response designated for study by Skinner (24). It also has been characterized as *object-consistent*

response, as opposed to ***body-consistent*** response (**6**). Without implying that there is agreement on how such behavior can be explained, it can be stated that this is the kind of response most frequently studied in contemporary learning theory. We may quote Spence on this, speaking for the "S-R Association-Reinforcement" point of view with which he identifies Hull and himself.

> Both the stimulus and response are described at a molar, commonsense level, rather than in molecular, physiological units. Thus a response is described either in terms of the effects it produces in the environment, such as depressing the lever of a Skinner Box, or in terms of the changed spatial relations of the organism in its environment, for example, entering a blind alley, approaching the positive stimulus, and so on. No note is taken of the differences in the detailed movements or motor pattern of the activity. Thus all movements of the organism that result in the same environmental change are regarded as a single-response class (**25**, p. 247, footnote).

In the present discussion, we shall use the term "object-consistent" for such flexibly fitting responses. In contrast, a body-consistent or muscle-consistent response, controlled only in terms of body parameters, may fit an environment, but only on the basis of memory —it runs itself off without the *flexible fitting* which characterizes the object-consistent response.

From the considerations just reviewed, the present paper assumes two problems: ***First,*** it assumes that the problem of adaptive fit in the execution of habits (or in the execution of adaptive instincts for that matter) is a puzzle needing explanation, and one as vulnerable to teleological pseudo-explanation as were the problems of inherited organismic form and learning. Such a focus is not common in contemporary psychology, as the level of analysis chosen for data collection and theory avoids the problem. But the focus is shared by Guthrie, if by no others. ***Second,*** it undertakes to apply the selective survival model to such adaptive fit. This application proves difficult, and the task may seem gratuitously undertaken. Persistence seems justified by the success of the model for the other levels of adaptive fit, and by the absence of other explanatory frameworks for apparently teleological sequences of behavior. Application of the model to the object-consistent response requires that a random variation and selection process go on in carrying out each response, to account for the fit to environment which it shows. The existence of such a random trial-and-error process appears more readily if we examine first object-consistent responses on the part of organisms lacking distance receptors.

Blind object-consistent responses. Let us consider a blind person who has learned the task of sorting mixed machine parts into separate bins. The response of reaching is for the most part a body-consistent response, guided in terms of body orientation and memory. The final phase of the grasping, however, involves an observable blind, random groping, varying in extent depending upon the accuracy of the initial movement. Without the trial-and-error component, the object-consistent adaptive fit of the response would not be achieved. An example of a blind, fumbling, yet object-consistent response for a seeing person might be finding a cigarette lighter in a pocket. But even for a blind person's behavior, this random trial-and-error aspect of executing well-learned behavior may not be too apparent. In part, this is due to the fact that, in a simple stable environment, body-consistent responses may approach fit quite closely, leaving the random trial-and-error component to a minimum. Thus in walking on a well-known level floor, the blind person may

place each foot without hesitation, or visible trial and error. And on an uneven terrain, the random trial-and-error component may be safely limited to one dimension, and appear only as a hesitant feeling for the ground. But if we are to follow the logic of Ashby's presentation, and avoid any unexplained prescience on the part of the organism, the random trial-and-error component must be present in every object-consistent response.

Let us follow further our blind subject in his well-learned performance. He identifies the piece by a random scanning of its surface with his fingers. He places it without apparent hesitation in the correct bin, if the bins are large enough for a body-consistent response to be adequate. But if we observe him more closely, we note that he has searched out the correct bin with his left hand, and that this has given confidence and precision to the response of the right. Somehow, the random trial and error of the left hand has vicariously served for the right one. Similarly, he may in walking use a blind trial and error of cane movements to search out steps, walls, and doors, reducing the trial-and-error component in his walking.

Guided object-consistent responses. But most object-consistent responses have a smooth, accurate, guided quality which seems quite out of keeping with the prescribed random trial-and-error process. If the formal model for adaptive fit is to be retained at this level, the only resolution seems to be to locate the trial-and-error process in the function of the sensory organs. It is the burden of this paper that perception serves this function of trial-and-error exploration, substituting for the motor trial and error found in the blind object-consistent response.

It is probably easiest to accept this point of view for an organ of vicarious exploration like an insect's antenna, or the blind man's cane. The analogy of the radar screen as an aiming device is of help. The radar beams scan the sky in a blind sweep, blind in that it is not modified by any prior knowledge of the location of objects. When in this search a beam reflects from a plane, a gun is then appropriately aimed. The trial and error of a radar beam has substituted for a trial and error of expensive bullets. In a parallel way, a ship's radar vicariously explores the waterways, by a trial and error of radar beams learning the location of obstacles that might otherwise have been located by a trial and error of ship movements and collisions.

It is an easy transition from the radar model to the bat's supersonic echo location—in which sound waves emitted in all directions provide the substitute random trial-and-error process. Similarly, the lateral line organ of fishes seems to have the purpose of registering waves of water pressure change in such a fashion as to locate objects in terms of the echo of the fish's own swimming, and Pumphrey (23) has suggested the radar and echo-location analogy for this process.

The case for vision is most important, but cannot be made with the clarity and completion possible for the radar and echo-location examples, since the emitting process is missing. However, the notion of vision as a surrogate trial-and-error process seems not only required by the formal model but supported by other considerations. If in visual search the gross eye movements are not blindly searching, it is because other sources of information such as touch, memory, or hearing have been employed to narrow the range of search. Hebb (11) has assembled impressive data on the active searching movements that typically characterize the simplest seeing process, and his facts

belie the implicit notion of the passive, fixed-focus eye implicit in both Gestalt and conditioning theories. But even without temporally extended scanning, the eye in a single glance provides spatial information which can substitute for motor trial and error, which can lead to smooth, guided, object-consistent responses.

Can this be fitted into the model? Perhaps. Ashby's Homeostat under stress presents to its environment a lot of alternatives, from which one that fits is selected. Similarly, the radar beam presents in its ever-repeated scanning sweeps multiple alternative loci for reflection. In both of these the alternatives are temporally extended. The rods and the cones of a fixed-focus eye can be regarded as the simultaneous presentation of a myriad of alternative loci for possible excitation, blindly available in that their location or availability does not anticipate the location of objects, except as this glance has been preceded by other glances and other sources of information. We could build a radar device in this manner, so that instead of one scanning beam of varying direction, it had a million simultaneously operating beam emitters and receivers, all of fixed aim. The learning capacity of Ashby's Homeostat lies in the range of settings it can try. The learning capacity of the radar lies in the range of directions in which it sends its beams. The learning capacity of the eye lies in the range of possibilities which it makes simultaneously available to selective excitation. Thus even without the emitting mechanism of radar, major portions of the model seem appropriate. Vision can be seen as providing data about the spatial environment intersubstitutable with what might be learned by blind trial and error. It is to be understood similarly in a deterministic way, with no appeals to prescience. It retains the basic epistemology of trying a lot of things and seeing what works. Although the analogy is not complete nor elegant, let us explore its implications for some already established problems and points of view.

Relation to Other Viewpoints

Response guiding through feedback. Cybernetics has made a valuable contribution to learning theory through providing a thoroughly deterministic model for the treatment of purposive behavior. Among the learning theorists, Guthrie (10) and Mowrer (19) have explicitly recognized this role. Guthrie (10, p. 283) and Wiener (37) have applied the concept of feedback to the execution of object-consistent responses, with a solution similar to the present one, but differing in an important respect. They both accept the analogy of the automatic pilot and the steam-engine governor. Applied to the object-consistent response, this means that the response is guided by its effects, as fed back by kinesthetic and other evidences of outcome. Guthrie notes, too, the "trial-and-error components" of such cybernetic controls. Both Guthrie and Wiener emphasize the importance of visual feedback in this process. "As the cat reaches for the post either ready to bite or ready to claw, the movement is continuously corrected by vision and therefore may be executed from a small variety of stances or of distances from the post" (10, p. 283). Yet the emphasis here is on the *after*-feedback of the results of the motor movement, as in the blind object-consistent responses described above.

In contrast, for the notion of perception as substitute trial and error, the automatic pilot and steam-engine governor are not adequate analogies. Instead, a distinction is drawn between these, in which the main motor output

is guided by its own effect, and those other cybernetic devices in which the main motor output is guided by the results of a prior substitute output and feedback. Thus the radar controlled anti-aircraft missile is not guided by feedback from the projectile's location or outcome, but rather by the feedback from a prior output of electromagnetic waves. The movement of the radar-guided ship is not controlled by feedback of the ship's contacts and collisions with other objects, but rather by the feedback of the contacts and collisions of the exploratory, substituting radar beam. Perception is seen as controlling guided distal responses in this same trial-in-advance way.

Lashley (15, p. 122) has argued against the theory of control of movement which "assumed that it is continued until stopped by sensations . . . which indicate that the limb has reached the desired position." He finds the speed of some controlled actions are such as to "force the conclusion that an effector mechanism can be preset or primed to discharge at any given intensity or for a given duration in independence of any sensory controls." While he is referring specifically to kinesthetic feedback, his remarks might be generalized to include visual feedback from motor movement. A study by Hess (12) on the pecking of chicks also seems relevant. The normal accurate pecking, clearly a guided object-consistent response, is made systematically inaccurate when the chicks wear distorting lenses. No amount of experience, motor feedback, or visual feedback from motor effects seems to correct this error. Instead the instinctive accurate pecking seems rigidly guided by prior visual search. The reader can perhaps convince himself of the distinction by noting with what considerable accuracy he can guide his hands to an object by first looking, and then closing his eyes before he reaches. In making this distinction, I should not like to imply that Guthrie would disagree with it. Guthrie's comments on this problem are an important step in resolving long-standing disagreements in learning theory toward which he has persistently called attention. To quote him:

> Advertent solutions transferred more readily to the post in a new position. The cat did not bite or claw the air in the old spot. The animal was more likely to look around and on seeing the post in a new position approach it and perform a guided act like biting or clawing. When escape had been inadvertent, this transfer to a new position did not take place and the cat repeatedly backed into the place where the post had been.
>
> Advertent solutions obviously have many of the qualities that interest Tolman (and should interest others). They can be described as expectancies or as perceptions of means-end relationships. It is our belief that in associating the act of reaching with the sight of the post, tendencies to reach out may through previous practice be conditioned on vision and visual orientation and serve as maintaining stimuli for a sustained reaction to the post which has the same trial-and-error components as has the automatic pilot of the plane or steamship. Reaching out and touching is a skill with much practice behind it, and it is also a behavior mode which exhibits control. When the telephone rings we ultimately reach the instrument even if our chair is in a new spot and we must follow a course which never before has been followed. We respond to the bell by rising and by being ready to grasp the telephone, perhaps by being set to say "Hello." Seen obstacles are avoided. That avoiding seen obstacles is based on past training is evident from recent operations on children for cataract in which seen obstacles are not avoided.
>
> In other words, association may result in acts as well as movements, and this is evident in cats as well as in men. The basic nature of the learning may be just as much an association of stimulus and response in an act that includes sustaining stimuli and cybernetic corrections as it was in Pavlov's salivary responses. The automatic pilot, the thermostat, the governor of the engine—all illustrate the fact that physical analogies are available in which by setting a control we govern the later behavior of a complicated machine. In animal behavior we have only to assume that

the setting, which can itself be a physical response, is itself subject to associative learning (10, pp. 283–284).

Vicarious trial and error. Our discussion of substitute trial and error is reminiscent of the concept of Vicarious Trial and Error (VTE) given currency by Muenzinger (21) and Tolman (31). The usage is not the same, however. Their focus was on abbreviated motor movements that substituted for complete motor acts in reactivating memories, or as an approximation of thought, or as a symptom of consciousness, rather than a process in which the substitute trial and error serves as a source of current information about the immediate environment, in equivalence to locomotor exploration.

In Muenzinger's first presentation (20, p. 204) he spoke of a type of behavior which "appeared to us to be a vestige of actual trial and error behavior," and indeed "vestigal trial and error" might have been the better label for what he had in mind, as Dennis (9) has pointed out. Yet in Muenzinger's most full presentation of the notion (21) he felt that, although the animal abbreviated his exploration by not actually going into the maze, the process was none the less a "testing out" of alternate courses of action. Tolman had still earlier, in 1926 and 1927 (28, 29) presented a similar notion of the role of a rat's "wiggling his nose from side to side and finally choosing" or "running back and forth" in reactivating, enhancing, or catalyzing discriminations and associations already partly learned. While Tolman later adopted Muenzinger's term for this type of behavior, he seems to have paid little attention to the literal connotations of the label and ends up seeing the behavior compelled by competing orientation vectors (31). Thus VTE, even if somewhat inappropriately labeled, has a well-defined research usage and a specialized theoretical role in speculations on the phylogeny of conscious thought which is still of interest (19). These connotations are distinct from, and more complicated than, the simple concept of substitute trial and error here being offered. In VTE most literally interpreted there is a vicarious search of a vicariously (through memory) represented environment. And while the process is instigated by the visible objects of the choice point, it is not conceived as a search of them, or a learning about them. In contrast, the notion of perception as substitute trial and error refers to visual search as a guide to motor response, as a substitute for blind exploratory locomation, limited to the visually accessible environment. This process may go on in the execution of well-learned habits or in the execution of instincts.

Perceptual versus mnemonic expectancies. While Tolman's use of VTE does not duplicate the notion here presented, his concept of "perceptual expectancies" as parallel in effect but distinct from "mnemonic expectancies" does duplicate it. Under these rubrics Tolman in 1932 (e.g., 30, pp. 96–97, 117–118, 134–139) clearly makes the point that vision can provide information and behavioral guides equivalent to those obtainable through motor exploration. He seems alone among learning theorists in thus recognizing perception as a knowledge process substitutable for motor trial and error. While the point receives major attention in *Purposive Behavior,* it has not since been elaborated by Tolman and his students, and has been omitted in more recent summaries of that position. The present argument differs from Tolman's mainly in the emphasis upon the basic role of blind trial and selective retention in all adaptation or knowledge processes, with the resulting effort to interpret perception as a random search process.

Disagreement as to "what is learned." In a previous review (6) attention has been called to the factual disagreement in available literature as to whether or not learned responses were characteristically object-consistent or body-consistent. In transposition experiments in which test conditions made possible the disentanglement of the two definitions, the learned responses were *object-consistent* in the majority of instances, but not in all. For example, in four experiments of Wickens (33, 34, 35, 36) 91 per cent of those who showed transfer had acquired an object-consistent response, 9 per cent a muscle-consistent response. Likewise, in the Guthrie and Horton (10) experiments both types of response occurred. Guthrie (10, p. 283) speaking for his own data, suggests that where the responses learned were visually guided (advertent), they showed object consistency rather than movement consistency under transposition experiments. Where the puzzle-box release movement was "inadvertent" the response consistency was in terms of stereotyped movements. Similarly, Dennis (8) found that normal rats took a short cut when a maze wall was removed, but that blinded rats traced the same path as before. It seems likely that the inconsistencies in Wickens' data might similarly be resolvable through the degree of visual involvement in the original learned habit.

Insight vs. blind trial and error in problem solving. The orientation of this paper inevitably leads to agreement with Lloyd Morgan (18) and Thorndike (27) as to the fundamental importance of trial and error in problem solving. And it would promote this principle as a basic epistemology. But the trial and error will only be "blind" in situations where vision is of no use, or for animals without sight. Thus overt trial-and-error locomotion will not always be present for animals with distance receptors. In many instances, particularly under those conditions that Köhler (14) has employed, the trial and error is handled by visual search. In others, such as in Thorndike's problems, such search is useless, and overt trial and error is all that is available. In still other situations, in which transparent fences or walls are employed, the solutions offered by visual search will be misleading, and will postpone adequate solutions until the blindly haphazard responses characteristic of frustration take over. In these terms, the data of Köhler (14), Thorndike (27), Adams (1), and Guthrie and Horton (10) do not seem so far apart. Certainly in Köhler's original accounts, there is ample evidence of trial-and-error processes, of fortuitous rather than logically deduced solutions. And in Tolman's insight studies with rats (30, pp. 169–170) the insight that appeared with an elevated maze did not appear with a tunnel maze which limited the rats' perspective. These observations can be extended to Hilgard's (13, pp. 335–338) concept of "provisional try and correction" which he would substitute for random trial and error. Such selected and already quite adequate provisional tries will only be possible on the basis of prior learning or current perceptual trial and error. And if the try is incorrect, the blind animal will learn nothing of the direction of the error, whether too high or too low, too right or too left, and thus cannot correct but can only try again blind-randomly. An animal with vision, however, can note not only the inadequacy of the try, but often the direction of the error, and thus can demonstrate the intelligent provisional try and correction.

In this orientation, a distinction can be made between insight or problem-solving behavior involving substitute

trial and error in the perceptually immediate environment, and higher levels of "thought." Most instances of animal insight, or non-overt trial and error, are limited to relationships in the perceptually immediate environment. But certainly there are higher levels of ideation, in which the solution involves representations of more than perceptually present objects. It is not clear that this latter level has been demonstrated in animals.

Asch (2) has clearly and sympathetically presented the analogy between natural selection and trial-and-error learning, but nonetheless finds trial-and-error learning inapplicable to human beings, at least. While the present point of view makes trial-and-error learning the basic process and the final resort, it would result in some criticisms of learning theories similar to Asch's. Trial-and-error and conditioning doctrines are most appropriate to blind organisms. These theories as developed so far neglect the role of the perceptual organs as distance receptors and sources of information about spatial relationships in the immediate environment. A single photosensitive cell could substitute for the eye as conceived in the theory of most learning experiments, or a single acoustic switch for the ear. This neglect has been justified from a "first things first" approach to the development of science, and by a scrupulous adherence to deterministic explanation. But the neglect must be eventually corrected. The radar analogy should help in expanding deterministic learning theory into the behavior of organisms with fully developed visual systems.

The role of distance receptors in locomotor automata. In the view of the present writer, psychology stands to gain much from the experimental construction of automata which attempt to imitate life. Among those scientists who have essayed this, none has had a clearer perspective on the important problems of psychology, nor has been more meticulous in the logical development of requirements and solutions than has Ashby. His has been a magnificent achievement, and one to which the present writer feels greatly indebted. Yet Homeostat does not seem to imitate life as successfully as Walter's (32) mechanical tortoise or Berkeley's (4) mechanical squirrel. Ashby's Homeostat does not locomote in the physical world. It stays put, and adjusts to the impingements of a very special environment. Nor has it the ability to recognize segments of the environment, and store unused response patterns for use when these recur. As Walter has commented, in imitating life it is more plant than animal. In contrast, the mechanical tortoise and squirrel locomote purposefully in our world of objects, using both locomotor trial and error and scanning photoreceptors. They appear life-like to the naive observer.

In the particular physical world in which we live and in which evolution took place, objects that are impervious to locomotion in general also reflect or diffuse certain electromagnetic waves. This persistent ecological condition over the eons has made possible the development of organisms able to anticipate the presence and location of solid objects through a substitute exploration based upon the opaqueness of the object to light. In the evolutionary perspective, vision is based upon this environmental correlation, as is our recent development of radar. The austere logical environment of Homeostat lacks such fortunate coincidences. No environmental parameters are provided which are correlated to other environmental parameters so highly that search of one can substitute for trial and error in contact with the other.

Is vision, or some other highly precise distance and space receptor, a pre-

requisite for locomotor object consistency on the part of animals or machines? No logical derivation is on hand to prove that this so. Our concept and illustrations of blind object-consistent responses argue against it. And yet organisms and machines achieving locomotor object consistency without a substitute exploratory mechanism are rarities. Blind individual animals do occur in seeing species and whole species of blind cave fish and moles exist. Yet rather than being the evolutionary predecessors of seeing forms, these represent the regressive evolution of forms that attained their complexity with vision. Certainly vision or echo location is the usual concomitant of locomotor forms of life, and a probable essential in a locomotor automaton.

SUMMARY

Selective survival among random variations is taken as a general paradigm for instances of organismic fit to environment. Darwinian theory of natural selection applies the model to the fit between the inherited characteristics of organisms and the opportunities provided by their habitats. Trial-and-error doctrines apply the model to learned fit between organismic response and environment. Ashby and Pringle have independently noted the formal parallel between evolution and learning.

Attention is called to a third level of organismic fit to environment, in the adaptive responses employed in the flexible execution of well-learned habits. For blind organisms, the trial-and-error component in the carrying out of a habit may be obvious; but for organisms with distance receptors, the smoothly guided yet flexible character of the execution of learned responses seems quite out of keeping with the random variation required by the model. An effort is made to resolve this incongruity by characterizing perceptual processes as substitute trial and error, containing a search component which takes the place of blind overt motor movements. The notion is a more primitive one than that of "vicarious trial and error." It seems relevant for the empirical inconsistencies in the problems of "what is learned" and "insight." As related to servo-mechanism models, the notion is to be clearly distinguished from the simple negative-feedback regulators, like thermostat or governor, in which the feedback comes from the outcome of the primary effector. But a suggestive parallel is available in complex servo-systems such as the radar-controlled guiding of ship or projectile, in which a blindly emitted beam is selectively reflected, and is used to substitute for a trial and error of ship movements or projectiles.

REFERENCES

1. ADAMS, D. K. Experimental studies of adaptive behavior in cats. *Comp. psychol. Monogr.*, 1929, **6**, No. 1.
2. ASCH, S. E. *Social psychology.* New York: Prentice-Hall, 1952.
3. ASHBY, W. R. *Design for a brain.* New York: Wiley, 1952.
4. BERKELEY, E. C. Light sensitive electronic beast. *Radio Electronics,* Dec., 1951, **23**, No. 3, 46–48.
5. BRUNSWIK, E. *The conceptual framework of psychology.* Chicago: Univer. of Chicago Press, 1952. (*Int. Encycl. unified Sci.*, V. 1, No. 10.)
6. CAMPBELL, D. T. Operational delineation of "what is learned" via the transposition experiment. *Psychol. Rev.*, 1954, **61**, 167–174.
7. CAMPBELL, D. T. Adaptive behavior from random response. *Behavioral Sci.*, 1956, **1**, 105–110.
8. DENNIS, W. The sensory control of the white rat in the maze habit. *J. genet. Psychol.*, 1929, **36**, 59–89.
9. DENNIS, W., & RUSSELL, R. W. Comments on recent studies of VTE. *J. genet. Psychol.*, 1939, **54**, 217–221.
10. GUTHRIE, E. R. *The psychology of learning.* (Rev. Ed.) New York: Harper, 1952.
11. HEBB, D. O. *The organization of behavior.* New York: Wiley, 1949.

12. HESS, E. H. Maturation and learning in the development of pecking accuracy in chicks. *Amer. Psychologist*, 1953, **8**, 367. (Abstract)
13. HILGARD, E. L. *Theories of learning*. New York: Appleton-Century-Crofts, 1948.
14. KÖHLER, W. *The mentality of apes*. New York: Harcourt, Brace, 1925.
15. LASHLEY, K. S. The problem of serial order in behavior. In L. A. Jeffress (Ed.), *Cerebral mechanisms in behavior*. New York: Wiley, 1951. Pp. 112–136.
16. LEWIN, K. *Field theory in the social sciences*. New York: Harper, 1951.
17. MACFARLANE, D. A. The role of kinesthesis in maze learning. *Univer. Calif. Publ. Psychol.*, 1930, **4**, 277–305.
18. MORGAN, C. L. *Introduction to comparative psychology*. London: Scott, 1894.
19. MOWRER, O. H. Ego psychology, cybernetics, and learning theory. In D. K. Adams *et al.*, *Learning theory, personality theory, and clinical research*. New York: Wiley, 1954. Pp. 81–90.
20. MUENZINGER, K. F., & GENTRY, E. Tone discrimination in white rats. *J. comp. Psychol.*, 1931, **12**, 195–206.
21. MUENZINGER, K. F. Vicarious trial and error at a point of choice. I. A general survey of its relation to learning efficiency. *J. genet. Psychol.*, 1938, **53**, 75–86.
22. PRINGLE, J. W. S. On the parallel between learning and evolution. *Behaviour*, 1951, **3**, 175–215.
23. PUMPHREY, R. J. Hearing. In *Symposia of the Society for Experimental Biology, IV. Physiological mechanisms in animal behavior*. New York: Academic Press, 1950. Pp. 1–18.
24. SKINNER, B. F. *The behavior of organisms*. New York: Appleton-Century, 1938.
25. SPENCE, K. W. Theoretical interpretations of learning. In C. P. Stone (Ed.), *Comparative psychology*. (3rd Ed.) New York: Prentice-Hall, 1951, Ch. 8.
26. SPIEGELMAN, S. Differentiation as the controlled production of unique enzymatic patterns. In *Symposia of the Society for Experimental Biology, II. Growth in relation to differentiation and morphogenesis*. New York: Academic Press, 1948.
27. THORNDIKE, E. L. Animal intelligence: an experimental study of the associative processes in animals. *Psychol. Rev. Monogr. Suppl.*, 1898, **2**, No. 8.
28. TOLMAN, E. C. A behavioristic theory of ideas. *Psychol. Rev.*, 1926, **33**, 352–369.
29. TOLMAN, E. C. A behaviorist's definition of consciousness. *Psychol. Rev.*, 1927, **34**, 433–439.
30. TOLMAN, E. C. *Purposive behavior in animals and men*. New York: Century, 1932.
31. TOLMAN, E. C. Prediction of vicarious trial and error by means of the schematic sowbug. *Psychol. Rev.*, 1939, **46**, 318–336.
32. WALTER, W. G. *The living brain*. New York: Norton, 1953.
33. WICKENS, D. D. The transference of conditioned excitation and conditioned inhibition from one muscle group to the antagonistic muscle group. *J. exp. Psychol.*, 1938, **22**, 101–123.
34. WICKENS, D. D. The simultaneous transfer of conditioned excitation and conditioned inhibition. *J. exp. Psychol.*, 1939, **24**, 332–338.
35. WICKENS, D. D. Studies of response generalization in conditioning. I. Stimulus generalization during response generalization. *J. exp. Psychol.*, 1943, **33**, 221–227.
36. WICKENS, D. D. Stimulus identity as related to response specificity and response generalization. *J. exp. Psychol.*, 1948, **38**, 389–394.
37. WIENER, N. *Cybernetics*. New York: Wiley, 1948.
38. YOUNG, J. Z. *Doubt and certainty in science*. Cambridge, England: Cambridge Univer. Press, 1950.
39. YOUNG, J. Z. Memory, heredity and information. In J. Huxley (Ed.), *Evolution as a process*. London: Allen & Unwin, 1954.

(Received January 25, 1956)

Psychological Review
Vol. 63, No. 5, 1956

SUBCEPTION: FACT OR ARTIFACT? A REPLY TO ERIKSEN

RICHARD S. LAZARUS

Clark University

In the January, 1956 issue of this Journal, an interesting critique was offered (2) of the subception experiment reported some years ago by Lazarus and McCleary (4). This is not the first such article which has appeared on the subject. Indeed, there have been several (1, 3, 7), and there have been a few reports of experiments which duplicate and extend the findings of the original study (5, 8). The recent article by Eriksen calls for some examination of the arguments.

In the Lazarus and McCleary experiment, tachistoscopically exposed nonsense syllables which had been associated with painful electric shocks stimulated larger GSR's than nonshocked syllables even when the exposure speeds were too rapid for the subject to identify them correctly. This effect was termed "subception," and a process of autonomic discrimination in the absence of the ability to report conscious recognition was suggested.

Eriksen analyzes the data in the subception study as an instance of partial correlation. He suggests that the subception effect involves a positive correlation between the stimuli and the GSR's and between the stimuli and the verbal responses. The fact that a correlation remains between the stimuli and the GSR even when the relationship between the stimuli and the verbal responses is partialed out is the subception effect itself.

The kernel of Eriksen's argument lies in his analysis of the number of categories in the two response systems. He points out that the subject is restricted in his verbal response at each presentation of the stimulus to a choice of one of ten syllables. However, no such restriction applies to the GSR system. Thus, if a subject correctly identified some of the letters of the stimulus, but not enough to perceive the entire stimulus correctly, his verbal response might not reflect the partial discrimination which had occurred although the information could trigger the GSR. Such conditions would increase the likelihood of a partial correlation which would favor GSR discrimination in the absence (partialling out) of verbal discrimination.

Eriksen's argument leads him to discard the subception process as an artifact of the statistical conditions imposed by the experimenters. He analyzes these conditions as though there were no doubt that the GSR system is indeed continuous and unrestricted in the functional sense. In his discussion there is the hidden assumption that the subject can really use all of the categories of the GSR theoretically available to him. The measurement which we obtain from the GSR may be continuous, but we do not know the nature of the process underlying the GSR, and the extent to which it is actually restricted or unlimited. The continuity of the measured response may be determined by a discontinuous process. The argument then becomes a matter of which assumptions one wishes to make about the process underlying the measurements themselves.

In his article, Eriksen makes the following statement about the Lazarus and McCleary formulation of subception:

> These authors do not clarify the nature of this process (subception), but their definition and their interpretation of their data imply

that the autonomic nervous system, under certain circumstances at least, is capable of making more accurate discriminations among stimuli than exist in the individual's awareness (2, p. 74).

The same kind of statement is made in the critique of subception by Murdock (7, p. 571): "Perhaps the proper conclusion to be drawn from their study [Lazarus and McCleary's] and that of McGinnies as well is simply that the GSR is more accurate and/or sensitive than verbal reports as a measure of recognition."

It is not at all necessary to assume that the physiological response system of the organism is a more precise mirror of the physical stimuli than the verbal response system. A perfectly logical alternative is that the autonomic response system reflects the presence or absence of danger—that is, the shock or nonshock consequences of the stimulus—even though the level of discrimination is not sharp enough to identify the specific components of the stimulus. In this event, the process does not involve a one-to-one correspondence of the autonomic response with the stimulus, but a categorical one. Such a relationship may be based upon the direct response of the organism to the stimulus, or may be mediated by a process of inference in which the nature of the stimulus is built up from the stimulus elements. The latter is exactly what Bricker and Chapanis seem to be suggesting with respect to the verbal response system. The subject uses partial information about the stimulus to make an inference which may turn out to be wrong and not even verbalizable.

Thus, a number of conceptual alternatives arise in our attempts to interpret the data of subception. (*a*) The autonomic response system may mirror directly the stimulus properties, independent of verbal activity. (*b*) Alternatively, it may reflect a process of inference about the affect-laden aspect of the stimulus. In the latter case, the autonomic response might or might not depend upon elements of information being articulated verbally. There might be enough information getting through to the verbal sphere to identify the danger, but not enough to perfectly articulate the total stimulus. We are sometimes frightened without being able, at the moment, to indicate verbally what we are frightened of.

There is a most interesting aspect to Eriksen's proposal which is not carried to its logical end. Eriksen implies that, in order truly to demonstrate subception, the number of verbal response categories must be equivalent to the number of GSR categories, or, more pertinently, that the number of verbal categories be no less than the number of discriminations that can be made to the stimuli. He is thus suggesting an experiment which would reflect such a condition—with the implication, it seems, that the subception effect could not then be demonstrated.

Eriksen's argument is based on a statistical rather than a psychological analysis of the problem. It could be also argued that restriction in one response system (verbal) in contrast to the other (autonomic) is quite close to the conditions prevailing in actual life, and that what we refer to as "lack of awareness" can be dependent upon (although not necessarily identical with) just such restriction in the verbal processes. Let me elaborate this point because I think that it gets to the heart of Eriksen's position.

The properties of language are such that it always represents a discrete distribution of responses in relation to the stimulation from the physical world, while the physiological response has no such restriction. Consider, for example, the physical quality of color in relation to the verbal response system

which we use to describe it, and the sensory physiology of the organism which also responds to it. We define color in terms of wavelengths of light which are distributed in a continuous pattern. But we do not have language to identify each discriminable difference in wavelength. There are large gaps in the physical stimulus as it is reflected in the verbal categories, red, yellow, green, etc., even when we made much finer gradations such as chartreuse or aquamarine. This same point may be made for all verbal categories, whether they are applied to simple physical events or to complex social events. Furthermore, anthropological data tell us that additional categories of verbal response can be acquired which emphasize certain stimulus aspects of the physical and social world that are functional for one culture and not another.

By extension, the inability of a person to communicate (and hence to give evidence of being aware of certain stimuli) arises partially from the restrictedness of his language. By this postulate, therefore, discrimination without awareness (and perhaps many processes which are called unconscious) is a function of the restriction of one of the response systems involved, the verbal one. If one makes the dynamic assumptions of the clinician, then the response system is restricted through the operation of certain hypothetical ego processes. In terms of reinforcement learning theory the verbal response system is limited in accordance with rewards and punishments. But the inability of the person to categorize verbally the events to which he is responding can still be viewed as depending upon the condition of verbal response restriction.

From this standpoint, subception is an appropriate empirical model of what actually happens in nature instead of being an artifact of the conditions imposed upon the subject by Lazarus and McCleary. It is quite reasonable and compatible with experience to propose that the individual responds discriminatively to stimuli prior to his being able to articulate the stimulus verbally, although it is likely that discrimination at the autonomic level is quite a different process from that which we usually refer to as perception. The latter involves verbalization and hence awareness. Surely such discrimination takes place in children long before the child can describe his experience through language. How the organism can discriminate the stimulus subverbally is as much an unsolved problem, of course, as how the discrimination can be articulated in speech.

I do not know whether it can be maintained that the autonomic system is a better reflector of the stimulus than the verbal response system. However, organisms unable to use language often seem able to discriminate some stimuli better than man, and the absence of a cerebrum may even result in facilitated discrimination. For example, there are few instances of pigeons being hit by automobiles, although in our large cities they are certainly exposed to this danger. Moreover, a decerebrate frog can apparently catch flies better as a result of his decerebration. It is clear that verbal responses are not necessary for extremely sensitive discriminations, although they surely function importantly in what we call perception. Why reify verbal responses in psychology as though their absence makes discriminative behavior impossible? The subception effect suggests not that autonomic discrimination is better than verbal perception, but that it can be prior or responsive to aspects of the stimulus which are not verbally articulated.

Of the papers which have dealt with the subception effect, none have questioned the existence of the phenome-

non. The issue has been how this effect should be regarded. I have discussed here a number of issues which might lead to experiments pertinent to the clarification of the nature of subception. Let us take the more important theoretical possibilities and consider briefly some of the kinds of data which might be useful in each case.

1. Does the subception effect depend upon the limitation in the number of categories in the verbal response system? This is Eriksen's statistical proposal, and experimentally it would involve extending the number of verbal response categories to equal more nearly the number of categories discriminable in the stimulus. Even if the subception effect were not found under these conditions, however, the question would remain as to whether these conditions actually apply in nature. If there is greater isomorphism between the stimulus object and the GSR than between the stimulus object and the verbal response system, then Eriksen is merely affirming that the subception experiment takes advantage of the way in which man is actually constructed to show autonomic discrimination in the absence of verbal articulation.

2. Does the GSR mirror the stimulus object directly or does it result in discrimination through some global process of inference about the threat or nonthreat meaning of the stimulus? One possibility here is to "condition" each syllable to different intensities of shock. If the GSR mirrors the object fairly closely, then it will respond differentially to each syllable. If the discrimination is in terms of gross categories like danger versus no danger, then differentiation between GSR responses should not occur.

3. Does the GSR discrimination depend at all upon the ability of the subject to verbalize the nature of the stimulus? One approach to this question could require the subject to identify only the shock or nonshock nature of the stimulus rather than asking him to identify the specific syllable. If the subception effect were found here as well (GSR discrimination in the absence of verbal articulation of the shock or nonshock categories), then one might argue that the GSR discrimination does not depend upon verbal identification of the danger. This finding would not be crucial, however, since one could argue that partial information about the stimulus (e.g., combinations of letters) is used by the subject in making an inference from which the GSR discrimination is made, even though the subject could not correctly categorize at the verbal level. However, I believe that this latter problem can never be resolved since the elimination of this possibility cannot be made without also eliminating the variation between stimuli from which discriminations are made.

In regard to this third issue, one other experimental approach occurs to me. I am intrigued by the possibility that aphasic patients with loss of the ability to articulate visual stimuli verbally might nonetheless show GSR discrimination. Language in aphasia has recently been discussed by Werner (9) from a genetic frame of reference. In his microgenetic experiments Werner notes that "stimuli aroused 'feelings of word meanings,' inner experiences of the semantic sphere of the linguistic forms; these were apparently prior to any specific visual articulation of words." In reviewing other work on the problem, he cites the classic example of the subject who reported the word "cigar" as "smoke" which suggests that the correct word was sought by means of a meaning sphere rather than in terms of the visual articulation of word elements. It seems to me that the subception type of experiment could

be performed with aphasics, and perhaps even with mutes who cannot use words in articulating stimuli. Should autonomic discrimination be shown in such instances, it would be difficult to dismiss the idea that discrimination of meanings can take place prior to or independent of verbal articulation of the stimulus object.

I am aware that the subception experiment of Lazarus and McCleary does not really prove the existence of autonomic discrimination without awareness any more than the Bricker and Chapanis experiment disproves it. No experiment really *proves* anything, although the experimentation is important in determining whether the empirical consequences of a theory are consistent with it. Psychologists find great, and I believe often fruitless, sport in examining data which arise out of theoretical frameworks which have real fertileness for generating hypotheses and in showing that these data can be explained, *post hoc*, by some other system. Of course it can. In this sense, Eriksen's argument misses the point, as do those of Bricker and Chapanis, of Murdock, and of Howes. The issues dealing with such postulates as perceptual defense (6) and the process of subception cannot be decided solely on the basis of laboratory experiments. These issues have to be cast in the much broader frame of one's entire conceptual view of human behavior. One question which we must consider is the extent to which a view is both fruitful of hypotheses and consistent with the observations of events in nature itself. The nature of the subception process remains a theoretical question. The conditions which define its operation certainly can be clarified by further experimentation.

REFERENCES

1. BRICKER, P. D., & CHAPANIS, A. Do incorrectly perceived tachistoscopically stimuli convey some information? *Psychol. Rev.*, 1953, **60**, 181–188.
2. ERIKSEN, C. W. Subception: fact or artifact? *Psychol. Rev.*, 1956, **63**, 74–80.
3. HOWES, D. A statistical theory of subception. *Psychol. Rev.*, 1954, **61**, 98–110.
4. LAZARUS, R. S., & MCCLEARY, R. A. Autonomic discrimination without awareness: a study of subception. *Psychol. Rev.*, 1951, **58**, 113–122.
5. LOWENFELD, J. Verbal inhibition in subception. Paper read at Eastern Psychol. Ass., Atlantic City, March, 1952.
6. MCGINNIES, E. Emotionality and perceptual defense. *Psychol. Rev.*, 1949, **56**, 244–251.
7. MURDOCK, B. B., JR. Perceptual defense and threshold measurements. *J. Pers.*, 1954, **22**, 565–571.
8. RUBENFELD, S., & GUTHRIE, G. M. Stimulus generalization and subception. Paper read at Eastern Psychol. Ass., Atlantic City, March, 1952.
9. WERNER, H. Microgenesis and aphasia. *J. abnorm. soc. Psychol.*, 1956, **52**, 347–353.

(Received March 1, 1956)

ERRATUM

In an obituary article on David Katz by R. B. MacLeod (this Journal, 1954, **61**, 1–4) the statement was made that "Out of his chicken yard came the well known *Hackgesetz*. . . ." Professor Schjelderup-Ebbe, whose name has always been associated with the *Hackgesetz*, has kindly pointed out that it was he and not Katz who initiated and conducted the experiments. Katz merely edited the manuscript for publication and wrote an appendix in which he discussed the broad implications of the research.[1] The writer is happy to correct any false impression that may have resulted from his carelessness.

[1] T. Schjelderup-Ebbe, Beiträge zur Sozialpsychologie des Haushuhns; D. Katz, Tierpsychologie und Soziologie des Menschen (Anhang), *Z. f. Psychol.*, 1922, **88**, 225–264.